ISOLATED GENOMIC POLYNUCLEOTIDE FRAGMENTS FROM CHROMOSOME 7

5 **PRIORITY CLAIM**

This application claims priority under 35 U.S.C. §119(e) to provisional application serial no. 60/234,422, filed September 21, 2000, the contents of which are incorporated herein by reference.

FIELD OF THE INVENTION

10 The invention is directed to isolated genomic polynucleotide fragments that encode human SNARE YKT6, human liver glucokinase, human adipocyte enhancer binding protein (AEBP1) and DNA directed 50kD regulatory subunit (POLD2), vectors and hosts containing these fragments and fragments hybridizing to noncoding regions as well as antisense oligonucleotides to these fragments. The invention is further directed to methods of using these fragments to obtain SNARE YKT6, human liver glucokinase,

15 AEBP1 protein and POLD2 and to diagnose, treat, prevent and/or ameliorate a pathological disorder.

BACKGROUND OF THE INVENTION

Chromosome 7 contains genes encoding, for example, epidermal growth factor receptor, collagen-1-Alpha-1-chain, SNARE YKT6, human liver glucokinase, human adipocyte enhancer binding protein and

20 DNA polymerase delta small subunit (POLD2). SNARE YKT6, human liver glucokinase, human adipocyte enhancer binding protein and DNA polymerase delta small subunit (POLD2) are discussed in further detail below.

SNARE YKT6

25 SNARE YKT6, a substrate for prenylation, is essential for vesicle-associated endoplasmic reticulum-Golgi transport (McNew, J.A. et al. J. Biol. Chem. 272, 17776-17783, 1997). It has been found that depletion of this function stops cell growth and manifests a transport block at the endoplasmic reticulum level.

30 **Human Liver Glucokinase**

Human liver glucokinase (ATP:D-hexose 6-phosphotransferase) is thought to play a major role in glucose sensing in pancreatic islet beta cells (Tanizawa et al., 1992, Mol. Endocrinol. 6:1070-1081) and in the liver. Glucokinase defects have been observed in patients with noninsulin-dependent diabetes mellitus (NIDDM) patients.   Mutations in the human liver glucokinase gene are thought to play a role in the early

35 onset of NIDDM. The gene has been shown by Southern Blotting to exist as a single copy on

chromosome 7. It was further found to contain 10 exons including one exon expressed in islet beta cells and the other expressed in liver.

Human Adipocyte Enhancer Binding Pr tein

5   The adipocyte-enhancer binding protein (AEBP1) is a transcriptional repressor having carboxypeptidase B-like activity which binds to a regulatory sequence (adipocyte enhancer 1, AE-1) located in the proximal promoter region of the adipose P2 (aP2) gene, which encodes the adipocyte fatty acid binding protein (Muise et al., 1999, Biochem. J. 343:341-345). B-like carboxypeptidases remove C-terminal arginine and lysine residues and participate in the release of active peptides, such as insulin, alter

10   receptor specificity for polypeptides and terminate polypeptide activity (Skidgel, 1988, Trends Pharmacol. Sci. 9:299-304). For example, they are thought to be involved in the onset of obesity (Naggert et al., 1995, Nat. Genet. 10:1335-1342). It has been reported that obese and hyperglycemic mice homozygous for the *fat* mutation contain a mutation in the CP-E gene.

      Full length cDNA clones encoding AEBP1 have been isolated from human osteoblast and adipose

15   tissue (Ohno et al., 1996, Biochem. Biophys Res. Commun. 228:411-414). Two forms have been found to exist due to alternative splicing. This gene appears to play a significant role in regulating adipogenesis. In addition to playing a role in obesity, adipogenesis may play a role in ostopenic disorders. It has been postulated that adipogenesis inhibitors may be used to treat osteopenic disorders (Nuttal et al., 2000, Bone 27:177-184).

20

DNA Polymerase Delta Small Subunit (POLD2)

      DNA polymerase delta core is a heterodimeric enzyme with a catalytic subunit of 125 kD and a second subunit of 50 kD and is an essential enzyme for DNA replication and DNA repair (Zhang et al., 1995, Genomics 29:179-186). cDNAs encoding the small subunit have been cloned and sequenced. The

25   gene for the small subunit has been localized to human chromosome 7 via PCR analysis of a panel of human-hamster hybrid cell lines. However, the genomic DNA has not been isolated and the exact location on chromosome 7 has not been determined.

OBJECTS OF THE INVENTION

30   Although cDNAs encoding the above-disclosed proteins have been isolated, their location on chromosome 7 has not been determined. Furthermore, genomic DNA encoding these polypeptides have not been isolated. Noncoding sequences can play a significant role in regulating the expression of polypeptides as well as the processing of RNA encoding these polypeptides.

      There is clearly a need for obtaining genomic polynucleotide sequences encoding these

35   polypeptides. Therefore, it is an object of the invention to isolate such genomic polynucleotide sequences.

SUMMARY OF THE INVENTION

The invention is directed to an isolated genomic polynucleotide, said polynucleotide obtainable from human chromosome 7 having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:

5   (a)   a polynucleotide encoding a polypeptide selected from the group consisting of human SNARE YKT6 depicted in SEQ ID NO:1, human liver glucokinase depicted in SEQ ID NO:2, human adipocyte enhancer binding protein 1 (AEBP1) depicted in SEQ ID NO:3 and DNA directed 50kD regulatory subunit (POLD2) depicted in SEQ ID NO:4;

        (b)   a polynucleotide selected from the group consisting of SEQ ID NO:5 which encodes

10 human SNARE YKT6 depicted in SEQ ID NO:1, SEQ ID NO:6 which encodes human liver glucokinase depicted in SEQ ID NO:2, SEQ ID NO:7 which encodes human adipocyte enhancer binding protein 1 depicted in SEQ ID NO:3 and SEQ ID NO:8 which encodes DNA directed 50kD regulatory subunit (POLD2) depicted in SEQ ID NO:4;

        (c)   a polynucleotide which is a variant of SEQ ID NOS:5, 6, 7, or 8;

15   (d)   a polynucleotide which is an allelic variant of SEQ ID NOS:5, 6, 7, or 8;

        (e)   a polynucleotide which encodes a variant of SEQ ID NOS:1,2, 3, or 4;

        (f)   a polynucleotide which hybridizes to any one of the polynucleotides specified in (a)-(e);

        (g)   a polynucleotide that is a reverse complement to the polynucleotides specified in (a)-(f) and

20   (h)   containing at least 10 transcription factor binding sites selected from the group consisting of AP1FJ-Q2, AP1-C, AP1-Q2, AP1-Q4, AP4-Q5, AP4-Q6, ARNT-01, CEBP-01, CETS1P54-01, CREL-01, DELTAEF1-01, FREAC7-01, GATA1-02, GATA1-03, GATA1-04, GATA1-06, GATA2-02, GATA3-02, GATA-C, GC-01, GFII-01, HFH2-01, HFH3-01, HFH8-01, IK2-01, LMO2COM-01, LMO2COM-02, LYF1-01, MAX-01, NKX25-01, NMYC-01, S8-01, SOX5-01, SP1-Q6, SAEBP1-01,

25 SRV-02, STAT-01, TATA-01, TCF11-01, USF-01, USF-C and USF-Q6

as well as nucleic acid constructs, expression vectors and host cells containing these polynucleotide sequences.

The polynucleotides of the present invention may be used for the manufacture of a gene therapy for the prevention, treatment or amelioration of a medical condition by adding an amount of a composition

30 comprising said polynucleotide effective to prevent, treat or ameliorate said medical condition.

The invention is further directed to obtaining these polypeptides by

(a) culturing host cells comprising these sequences under conditions that provide for the expression of said polypeptide and

(b) recovering said expressed polypeptide.

35   The polypeptides obtained may be used to produce antibodies by

(a) optionally conjugating said polypeptide to a carrier protein;

3

(b) immunizing a host animal with said polypeptide or peptide-carrier protein conjugate of step (b) with an adjuvant and

(c) obtaining antibody from said immunized host animal.

The invention is further directed to polynucleotides that hybridize to noncoding regions of said

5 polynucleotide sequences as well as antisense oligonucleotides to these polynucleotides as well as antisense mimetics. The antisense oligonucleotides or mimetics may be used for the manufacture of a medicament for prevention, treatment or amelioration of a medical condition. The invention is further directed to kits comprising these polynucleotides and kits comprising these antisense oligonucleotides or mimetics.

10 In a specific embodiment, the noncoding regions are transcription regulatory regions. The transcription regulatory regions may be used to produce a heterologous peptide by expressing in a host cell, said transcription regulatory region operably linked to a polynucleotide encoding the heterologous polypeptide and recovering the expressed heterologous polypeptide.

The polynucleotides of the present invention may be used to diagnose a pathological condition in a

15 subject comprising

(a) determining the presence or absence of a mutation in the polynucleotides of the present invention and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or absence of said mutation.

20

DETAILED DESCRIPTION OF THE INVENTION

The invention is directed to isolated genomic polynucleotide fragments that encode human SNARE YKT6, human liver glucokinase, human adipocyte enhancer binding protein and DNA directed 50kD regulatory subunit (POLD2), which in a specific embodiment are the SNARE YKT6, human liver

25 glucokinase, human adipocyte enhancer binding protein and DNA directed 50kD regulatory subunit (POLD2) genes, as well as vectors and hosts containing these fragments and polynucleotide fragments hybridizing to noncoding regions, as well as antisense oligonucleotides to these fragments.

As defined herein, a "gene" is the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region, as well as intervening sequences (introns)

30 between individual coding segments (exons).

As defined herein "isolated" refers to material removed from its original environment and is thus altered "by the hand of man" from its natural state. An isolated polynucleotide can be part of a vector, a composition of matter or could be contained within a cell as long as the cell is not the original environment of the polynucleotide.

35

4

The polynucleotides of the present invention may be in the form of RNA or in the form of DNA, which DNA includes genomic DNA and synthetic DNA. The DNA may be double-stranded or single-stranded and if single stranded may be the coding strand or non-coding strand. The human snare YKT6 polypeptide has the amino acid sequence depicted in SEQ ID NO:1:

KLYSLSVLYKGEAKVVLLKAAYDVSSFSFFQRSSVQEFMTFTSQLIVERSSKGTRASVKEQDYLCH
VYVRNDSLAGVVIADNEYPSRVAFTLLEKVLDEFSKQVDRIDWPVGSPATIHYPALDGHLSRYQN
PREADPMTKVQAELDETKIILHNTMESLLERGEKLDDLVSKSEVLGTQSKAFYKTARKQNSCCAI
M

and is encoded by the genomic DNA sequence shown in SEQ ID NO:5:

CCAGACATAGGCAAGGCGCAAGGTGATACAGTAGGCAGCCACCATGGGGGGCCAGGAGGCTCC
AGCAGAGGCCACACAACCAGCCCAGAATCCAGGACAGAGAGCTGGAATGGAGACAGGGAAG
CCAGATACCAGGCCAGACTGGCCAGGTGCTACAGGCCTGTGGGCCAGGCCAGGCTTGGGGAC
TTCGTCCTGGGTGTGAAGGAGACAGGCACCCCTGAGGCCTTCCCTCTGCATCTCCAGCCCAAG
CTAAGCGCAAACTCTTAGGTTGGAGTAAGGAGTAACCCCCTGCCAAGTTTCTCCTGTCCTCAG
GCTCCACCCACCACCTATGCTGCCTGGCCCCATGGGGCACACGCTCAGGCCCAGCCTGGGAAA
GCAACTGCACCTGCCTGTGCTATGCTGGCCCTTCTCAGCCTCAATGCCCTCCTCCCTCCCCGAC
GCACCCTCGTGGCCCCCGCTGGGCCCCCTGATGCACCCTCATGTCTCCATGGCAACCTGCTCA
GAGTGTGGCCCTGCCCTTGGCTCCCCTCCACACCTGTGTCCCAGGCAGTGCCACGGCACTTTCC
TAAACAGAAGGATGGGCTTCAAAACAGTCCCAGACACTAAACACACCTGCATTTTGGGTCCA
AGTAACTTCTGACAAGACGAGTGCCCCTACACACTCTCAGTCCTATCCACTATGGGCAAGGAG
CCTGAAGGATCCCCCAGAACTGGCTAAAGCCCTCAGTCTCCTCCTCCACCCTGAGCACCTTCA
CGCGGCAGAGTGGCCCTGGATGTCAGCTTCTTGCTCCCCATGGTCTGCACCTGGACAGGTGCT
CTCAGGTGTGTGGGTGGGCAGGTGGCAGGTCCCAAGAGCCAGGTGCAAAGAATCTAGGCCAG
TGCCCACGAGTGCTGCAGTGTCTGTCCCCAGCATGGTATCTAGGGCTCCACTTGCCTATCAGCT
GTAATCGGAGGAGGCTTTCCAGGCCAGGCCTCCCCCAGGAAGGCTGCAGGCACTGCGGATCG
TGCGCCCTCACATGCATTATTCCTGAGGCCCTTCTGCAGATGCCATCAGGGCAGCAACTCTGA
TGAGGTATTAGGGCACAGCACACAGGGCTAAGCCACCCTGTACTGGGCCAAGCGCTACAGGC
AAAAAGGACACCACCGACGGGCATTTCATTCATCGCTTTTATTTTTATATATTTTTGAGAGGGA
GCCTCACTCTGTCGCCCAGGCTGGAGTGCAGTGGCGCGATCTTGGCTCACTGCAACTTCTCCCT
CCTGGGTTCAAGTGATTCTCCTGCCTCAGCCTCCCGAGTAGCTGAGATTACAGGTGCCCGCCA
CCATGCCCAGCTAACTTTTGTATTTTAGTAGACATGGGGTTTCACCATGTTGGTCAGGCTGGTC
TCGAACTCCCGACCTCAAATGATCTGCCTACCTCAGCCTCCCAAAGTGCTGGGATTACAGGCA
TGAGCCACTGCACCCGGCCCATTCATCACTTTTAAATAGCACCCTCTGAACAAAGCTCCCTGG
GCCACATGACCCTAAGGGTTACCCCATCCCACCCCAACCCAGGTCTGGCAGGTCCTCAGAACA

GGAAAAGCTGAGCACTGCCCAAGGCTGCTTGCTGGGCCAGTCAGAGAGGTCTCTGCCTTCCAG

GATCAGAAGTACAGGCTGAAAGCAGCCTTGGGCCCGCCTCCCTGGGAGGCTACAGAGGCTTC

AGAGGGTTCCCTGAACTCAAAACCAGATGTGAGACTTGAATTTGACTTACCCCTGGTTCACCT

CCCAACCAAAGCAGGGGTCAGCTTTGGCTCCTCCAGGAACCAGGAAGCTTCCAGGTACCCTGT

5 GGAGCCCCCTCTGCTCCTGAAAAGTTGCCACCTGTGCTTGGTGGGATGCCAGGTGGTCTCAGA

TTGACCCTGGGGTCAGCGGTGAGGGACAGGAAGCCTACAGCGGGATCAGGATGGGGATGGGG

CCTCCTGTCCCATGGCTCTGCAGCTATGAGGCAGCTTTCCTAGGGTGGGTCTCCTGGCTGCAGC

TAAGACCAGGCAACAGGATTCAGCAATGACAGGGCTTCTTCTACTCCAGGGCTCCCTCACCTG

GTTAACAGCAAAAAGAAAATACAGTTCCTGCTAGCAAGGTCTATAGAAGGAGGTGAAGGA

10 GTCAGGCCTGCAGCTACCTCTCCTGGACAGGAGCTGGTCAGGATAACTTGGACCCTTGCATGC

GGCAGGCCCACAGGCACACAGCATGAGGCCACTCTCTCCCCCGGGGGAAGGGCTTGGTGAAG

AAAGGATTCCCCTGAAGCACAAAGAAAGCACAGGACCACTGTGAAATTTCAAGACAACTTTA

TCCAGACAGGCGCCTCTCAAATAGAACACAGGGAAGTTAGGCAGCAGTTACTAAAATACAGT

CTCGCCAAATGATTTACAACAGAACACAACAGGAGCAGGGGATCTGTGGGTGGGGCTGGGCT

15 GGGCCCTCTATCTCACAGGGCCTGAGTCAAGCCAGCCCGCCCTGCAAGGCAGGGGCTGACCT

GCAAGCGGAGATCTCACTTCCTCTTACCCCAAATTCATACCTCCATTTTCCCCGCCCCCATCTC

TCCCCAGGGTCCTCAAGTGGGAAAGGGAGAGGTAGCATCCCTCGGATCCAGGCCCACTCCAC

TCCGTCTCCGGCACCAGTGGGCAGGCTGAGTCTGGGCCTCAAGGGGCCCTGGGCTTAGGGTAT

CTATGGCAGTAGGAAAATGACATGGACAGGCTCTTCAGGGGTAGGCTAAAGTCCTCTGGCCA

20 GCAGTACCCAGAGAAAATGGGCAGCAGCAGGTAAACCAGCCAGGAGGTGGAGTCCTCTGAAC

CCACAGCAGACCCCACCCTCCTGCCCAGCCCCTGCCCACATTGGGGGTCAGGACCACTGAGAC

TCTGGTCAGGACAGTGGGTGCTCTCAGCAGTGTGGCAAGCTCAGAGCAGAGCTCCCAAGGAC

CATACCACACTGGTTCAAAACCCATAGGTGACACCATCCCAGCAGAAGCTTCCATGGGTGCTG

GATCCCAGGGCTGCATCCTGAGCACAGGTGGGCAGACTGGAACATAACACTAGGACCCAAGG

25 GATCCAGAACATTTTAGGCCCATCTCCTGGGCTGCTCCAGCCTGTTGCCATGACTTGGGCAGT

GAGTGGGCCTCCTGCCAGGTGGCAGGGCACAGCTTAGACCAAACCCTTGGCCTCCCCCCTCTG

CAGCTACCTCTGACCAAGAAGGAACTAGCAAGCCTATGCTGGCAAGACCATAGGTGGGGTGC

TGGGAATCCTCGGGGCCGGCTGGCACCCACTCCTGGTGCTCAAGGGAGAGACCCACTTGTTCA

GATGCATAGGCCTCAGGCGGTTCAAGGCAGTCTTAGAGCCACAGAGTCAAATAAAAATCAAT

30 TTTGAGAGACCACAGCACCTGCTGCTTTGATCGTGATGTTCAAGGCAAGTTGCAAGTCAAGGC

AAGTGTCCCAGAGGCCCTGGGCAGCTGAGTGCACCTGTGTTTGATCTTCCCCTGATGATGGAC

ACTCCCAGCTGACCATCCAAACACCAGGAAACATCCCCCTTTCCTGGGCTCAGTTCCTAGTC

TACTTGCTGGTACGAACCCAACCCACACACTCCCCGCCCACAATGCAGCTCCTTCCAAATCCT

CCCACAAGCCACCTTTGTGGGACTTGGAAGCTGCTTAGGATGGGCCCTGCCCTCTGCGGGAAG

35 CCAATCCTAGCAGAAAGGTAAGCTAAACAACAGTCTCAGAATCTGAGACCCAGTGACT

6

GTTCCCCCCGCCCCAGGCCTTGGGCCTGAAGTGGGGGCCTGCCTGTGGCCTCTGTGGTGGGCT
CACTCCCACCCCCAACAGTGGCCCCAGGAGAGGCTTTCCCAAGAGTCTTCAAACTCCACCCAC
CCCAGCCCTAGCATCAGGGACTCCCCACCCCCCACTGGAGTGTTAATATCATTAATGTACAAA
TAAGATCCAAAGATATACCAAAGATCGAGAACAGCTGGCTCCGACCTCCCTCCCACAGAGC

5   CTTCCCAGGGTTAGCTGAAAAGAGCCCTTTGGCATCTACAGAAGCCAGTCGGAGTTTATGGT
TTCATTTGCCCAAAAATACACCTTTGGGGACCTCAAATTCTTTCCAAGAATCACTACCACACAT
ATGAATTTGAACATTCGCCACCCTTCCACCATCCATTTCTCGCAGGAACTTCAAAATAAAAAT
GGCCAGTCTGCCCCCACTCTGGCTCCTCGTCTATGGCTGTCTCTTCTTTTCCAGGGGCTGCAGT
TCTGATGTGAATGATGGTGCCATTCCAGCATTGGGCCTCTGGCAGGCTGCATCACATGATGGC

10 ACAGCATGAGTTTTGTTTCCGGGCCTTGGAAAAAAACAAAGAGGAGCTGAGAAGGAGGACTG
ACGAAGTAAGGGAAGCCCCAATCCTGGCAGGCGTGGCAGAGGGAGCTCCACAGGACACAGC
CAGGCAGAGAAACTAGCACTAGAACAGGGTGGGGGTGGAGGCCTTGAGGGAAGCTGTCCAC
AAGCAATTCCCATCACCAAGCACAAGGCGGGCCCCGGCTTCCAAAACTAGTCTGGGATCCTTT
TTCCTTTCTTTTCTCACACCCCATTAATGCTATCAAAAAGTGAGTAAAATTCCTACAGTTAGGC

15 CAGGTACAAACAAAGGACCAATAATACAAATGGGATTGGCAGAATATCTTAACTTTGCCCCA
CTCCTGTCTTCACACAATGCTATCTGACCACCACGGTGGTGTTTCTTCCTAGAAGATGGTCCTG
AGGACAACAGATGTGGTTCCCACTTGGGATGTGGTTTGTGGGGACCACTGTTGCCACCTTCTC
TCTTGCTTTCTGGTCACAGACTATCTTCCTAATCCCACCTAGCCATCTCCCTCCAATGTGCACA
TGAAAGCAAATGTGTGTGGACAGACCAAGTAAATTTGTCCCTATGACTATCCAACCATGGGCC

20 AACAGTGCCATCTCCACATAGGAAGACATGAGCACTGACCTGAGAGAAAGCGGCAGTCAGCA
GCACCCATCCTTGTCAATTAAATATTTTCTGTCAAAGGGAAATTAAAAGCTTAAGAACCTCTT
CAGGAAGGCTGAATTGCTTGCATCTTAAAGACTTATGTCTACTCAGCAGAAAGAGGAATAAG
ATTCAACAGTAAATCTCTGGTGATCAGAACTTGAACCAGCCTTCCTGGACTGGGAGTAGGAGT
TCAGAAATCAGCCAGAGCAGCAGAGGGCAGAGCAGAGGCAGGAGTGGAACAAGGCCTCGGC

25 CCGCATCGACTCCAACGGCGCCCAAGTGAACTGCCTCCAACCACCTGGGCCTGAGGCGCTCAC
CTTAGGCTCTTGCCGCACAAGGAATCATCCACCATGATTCAACAGTCTAAGAAAGACCCGTTC
ATAGTGGAGAGTGCCAGAAGCAGCAAGCTGCGACTGCTCTCTAGAGAGAACACCCAGGAGGC
AGCAGGTGCTGGGTACTCACAGTTTTATAGAAGGCTTTAGACTGTGTTCCCAGCACCTCGGAT
TTGGACACCAAGTCATCTAGCTTCTCACCTCGCTCTAACAGAGACTCCATGGTGTTGTGCTGG

30 ACAAAAAAGAAAAGAGAATCCAGCTCTGTTCAGTACGTGCCCTGACATGAGCCCCTCATATTT
CAGTCATGGGGGAAAGTGCCTTACCTGGGTTCCTCTCCAACACACACAAACTTCACCTCTAGG
TGTCGAGACTCGGTCCAAGAATAGTTACTGTCCAAGTGGATGGAACAGAACCTGGTGACATTC
CCGTGAAATCTAGAAGATCTAACTGGGATGTAGCAGACTTCCCAAAAAGCTGTCCCCAGCAC
AGGCTTAGATAACCAGCACTCCAGGAAAACTCATATATATATATACACACACATTTATATATA

35 CATTTGTGTGTGTGTGTGTGTGTGCACGCACATGTGCGTGTGCATGGAGCTTTGGAAAAAAGA

7

GTAGCTGGGCACTATATGATTGTACTGGGTTGGAGAGTGACCCACACCGCACCCCCCAACCCC

AACCGCATCCCAGAAATTAACATCCCCAGAATCTCTGAATGTGACCATATTTAGAAATAGGGT

CTTGGCAGATGTAACTAGTTAGGAAGAGGTAATACTGGATTAGGGTGGCATCTAATTCCATGA

CTGATGTCCTGGTAAGAAACGGAAACACACACAGAAGGTCACGTGACGGCAGAGGCAGA

5   GCCTGAAGTGATGCACCTCTAATCCAAGGAATGCCAAGGATGGCCAGCAGCCACCAGAGGCT

GGAGAGAGGCCTGGGACAGACACTCAGAGCCCCAAAAGACACCAGCCAGGCCCACAGAGCT

ATCTGTTAAAAGCAAATATTTGAGGGTTTCTGTTGACAGCAGCCACAGGAAACAAAAGGCGG

TGGGAAATGGCTATTGAGCACTTGATGTGAGGCAAGTCCAAACTGAGCAGCGCTCTGAGTAC

AGACACACCAGATTTCAGATGCAAACTCACACATGCTTCATTAGTAAGTTTTATACTGAAAAA

10   AAAACAAGTTTTATACCGATTACATGTTGGAAAAATTGTATTTGGATATACTGCGTTAAGTAA

AATATATAATTAAATTAAATTCTACCTATTTTCCTTTTATCATTTTAAAATATGGCTCCTAGAA

AATTCTAAGTTACACACATGCCCCAAATATATACCAGACAGCACTATGACAGAACATGTCCTG

CCTTCTAAATGGGCTATGTCCTAAATGTCATCACTACAAACTCTGACTTAGGAAATGAAAACA

CTGACCCCATGGGAAGGGGTCTAGAGATGGAGACCTCACAAGAGCCAGCAGCTCTGCTGCCA

15   GGGCCCTCAGGAAGCAGCAGCTCGCTTCTCTCCTCAGATGGCCACTGCTGCAGCAGCTAGATG

CACACATGAAGCGCCATAGAACAAGGAGCCAGCAAGAATGTCCTTCATCCCTACACACAGCT

GAGCGACTCAAATTTTTAACACAGAAAGTTAACTGATTCAGATATGCACACCAATCATCTAGA

TTTTACAACTGCAGCTAGATGAGGCTGGGTGAATAGGACTCATCCACTCCCCACCGTGGGGAG

AGGAGAAACAGCGGGTGTCCCAGGTGTCATGGTACTCAGACTAGGACTTGAGCAACAGAAAG

20   AGATGGCTTGAGGAGAAAACGGAGAAATGCCACCTAGGTGGTAAGAAAGCTCACAAGGTTTC

AAAAGACACAGATACCATGAGACTTTCACATCTATCGTTCATTCCAAAGCCACGTTATTTGGA

GTGCAGTCAGCACACCTGTGTTTGAAGCCCCTGGGATGCTTTTTATAAAATGCAGGTTCCCAG

GCTCCATCGCAGGCCAACAACTCCAACCCCAGGAGACGCTGATGTACACACTAAAGCTATGC

CTGTGTAAATGGTAAAGCTTTGTATGTGGGTTTCAATCCACTCCAGGTATCTATCAACTGCTGA

25   GCATGGTATAAACTAGGCACTGTATCATGAGCAGGATGGAAAGATGTCCCAGTGCTCATACG

CTGGTCAGGGAGACATGTAAACAAGCAGTGACAAAACTGTGACATCTGGTCAGAAAGGCCCA

ACCTTCAGGCGCCTGTGTGTGAGCTGGGCAAGAAAGGGTATAAGAGAGAACAGGGCCCAGTC

AGGAGACTGTGAGTTAGTTTGCACTTTATCCTGGGGCGGATCTGAGAGCTGCTGAAGGGTTCT

AAGTTGTGCAGATCAATGACTACTCTCTGGTGGACAGACTGGAGGTGAGCAGGAGGCAAGGG

30   GACCACTTAGAGGCAAAGGCTGTAAGAGAAAAACCTGAGAAAAACAGATAGCTGCTTACATT

CCACTTGTATGCAAAAATTTAAAAAAAAAGAGTTGAAGCAACAGTTACAAATCAGGAGATTT

CAGCTCAAAATGCAGGGTTCTGGCTCTTTTCAAAGGGGCCTATGTGACAACCCTGGGCCCATA

TTCCAGAAGCTGCCCTGTGGTCAGTGCACGGTGCTTCAATCTGTTCACCTTCAATGCAAACGCT

GCAAGGGGAGGCACCTGTGGGGTGTGGAGGCACCCGAAACCCTAACAAAGGCACCAGGGTG

35   GGAATCCAGGTCTTCAGAAGCCAAACCCTAGGAACCCAGTAAATGGTCAGACAGGCAGTAGC

CATGAGGAAGGGAGACTTGAGGGTTCCACTGGTTCCCAGCTTGGTCCCCTAGAAACAATGGGT
GCCATTAACCAAGAGAAGGGTATAGGAAAGACAGTCTGATGCCCGGGGTGGGGGAAGGGGT
GGGCAATCCCACTTGCTGGAGAGTGCCGTGGTTACTATTATATTAAAACGAGGATGGATCTGT
GCATGCCTGGCCAGTGGAAATCGCACCCCCGCCTCAGTTCTTGGGCTTGCTCTCCATCTTCCTG

5   CTTACCAGAATGATTTTGGTCTCATCTAGTTCGGCCTGCACTTTAGTCATGGGATCAGCTTCTC
GTGGGTTCTAGGAAAGAGTGAAAAATAATAAAGTCAGGACTGGAGTGGCTACCTGCAAACAA
AACCTAAAACTGAGGAAGCTGGACAAACTTTCACAGGTTAAAAACCACAGCCTGGGCCGGGC
ACAGTGGCTCACGCCTGTAATCCCAGCATTTTGGGAGGATGAGGCGGGTGGATCACCAGAGA
TCAAGAGTTCGAGACCAGCCTGACCAACATGGTGAAACCGTCTCTACTAAAAATACAAAAAT

10   TAGCCAGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT
CGCTTGAACCCAGGAGGTGCAGGTTGAGTGAGCCGAGATTGCGCCACTGCACTCCAGCCTGG
GAAACAGAGTGAGACTCCAACTCAAAAAACAAAAAACAAAAAAAAAAACCCACAGCCTGTT
TAACATGTAACAGAAACCCAAAGCCTGCCTAGAGCTTGGGTTCCCCGGTCTGAACGTAGATTC
TCTGTTTTCCAAACAGTAAGGCTTGAGAGAGGACACCAGCATCAGAAGCTGTCAGAAGTAATT

15   AGACCAGAACTATCAGGGCAGTTGGCTTTTTCAGTTTCACATGGATTCTGGGCCACATGGTGT
CTGCTGAAGCTTCCTTTAACCCTACCTGGTATCTACTGAGGTGACCATCCAGGGCTGGGTAAT
GGATTGTAGCAGGGGATCCTACTGGCCAGTCTATCCTGTCGACTTGCTTGGAGAATTCATCTA
GTACCTGCAAGACAAAGGAGACTCAACAAGCCTCCCACTGTGCACTCACCAGTGGTCTCAATG
ACAGGGCTTCACCCCTGAGCACCTCACCCTGAATGAGGCTCCTTGGCCTTCACAGCCCAGGAA

20   GGAGGAATGAGGGGGACATATAATGGCAACAGAGAAATCTAGGCTAAAGTTCTTTCCAAAT
TTTTATCATTAAAACATATCCTAAATATTCTGAGAATCAAAAGTATGCCCAGCCCGAGATGAA
CCTCACTTGGGGAGTAATAAAGGTATTTGAATTTTAAACTACAGATTTCCAGAAAAAAGGGGC
ACTGGTCCTCTAATTTTCCAAAGCAATTTTTTAAAAAAGAGAATTAGGTCCCCTAGATTTAAG
AAACCACCAGATTCCATGTGTTTGGAGGTATTTTGGTGCTCTGGGGTATAGGATGAAGCCTCT

25   GACTTCAAAGAGTTAATATTAGTAATTAGCACCGTACGCAAAAAAATTTAAAGAATGCTTAGG
TGCTAAGCTCTGTGGTGCAACTGACTGACATCAAGGTAGAGGGATGCAGCAACTGCAGGAGG
CAATGGGGAGAGTGAAGGCATTCAAGAGGGAGACTCCTTGAGCAGAAGCACAGGGGGCGAG
AACACAAGGCACAGCTGTCTCCGAGGGTCCCATCCCAGAGAATAGATGCTATGACTCAGTGG
CCTAGACCCAGCTCACATGAGGGACAGCACCGGGGAGGAAACCCATACAGGGATGCCAAATT

30   GTCTCTTGGGTTGCAGGGAAGGGGGCTGAAAAATGTGTTGACTTTGGACACATCATTTCATCC
CTTATGTCTCAGGGACTGCCATCAACCCCTGTCCCAGTCCATAAATGTGCCCATTCATCATCCA
AGTCCAGGAGAGGCAAATAAAAAACTCACCTTCTCCAGCAAGGTAAAGGCCACCCGGGATGG
GTATTCATTGTCAGCAATGACCACACCTGCAAGACTATCATTCCGGACGTAGACGTGGCACAG
ATAGTCTAAGGAGACAAGAGATCAGACACATGGATGCTGACATGAGGGCTTCAGACTTCTTTT

35   AATCCCCCCAAATCAAAGCATCCAATGTTAGGCCAAATGAAGCCACTCGGAAGCTCAATAGC

9

TCTGGGCAAGTCTTGTGGAGAGGCTTAGCAGCACAGCCCAATGGGCCACACACAGGAGCTTG

GCCCAACGCCTGCTTTAGGACCAGTAAATACCCAGAGGCCCAGTATGCAAAGCCAGGGCTTA

AAGAAACAGCCAGTGGTGCAGAAAACACACCCTTGACAACATGGCCCCAGGAGCATTTCCAA

GTGTATTCCTTAAGCTCGGGTCAGGCCAAGCTATATCTTAGGGATCTGGAGCCCTTGGGGCTC

5 TGTGCTGCTCCCAAACTTAGGGAACCCTGGACAAGCCAAGAGGCCTCTGCTTTCTTAAAAAAT

CTTTTCAGAGCAGCCAAAAGACAGGAAATTACCCCCCAGGGCCTCAGTCTTCCATATTATAGC

AACCTGCTGGGTTTGCTCCACTCTGGTGGGTGACTGGGAGTAGGGGGGTTAGTCTAGAAAAAG

ATTAGCTACTGCCAGCTAAGGCCTCCAGAGCACTGTGCTAAAATCCTCATATGATTGAAAGGT

ACAGTTGTACAGGTCTTCCGCAAAATATTCACAATCCACAGGATTGTTCATTTCCATCACTTTG

10 AAAGGATTCAGAGTTGATACAGCTAACCATATCCCCAAGGAAAGAGAAATGTAAGGATTACA

GCTTACAAATAAGAACCTTCTTGTCCTTAAGGATCTGACCCAGAAGATTCCAATGCTAAACAA

CAGAAAAACAAATAAAAGAGGAGGGAATGATGGTGAGCCCCTGAAATCAGAAAAGAGCAGA

GATAAATGAGAACAAGAATGAGGAGGAGGAAGAGGACAGGGGGTTGTCACCAATGCTCTCC

AGATTTTGTATACCATCCCCAATTAAGATTCAAACATGGGGTCAAAGTGCATACCCTCCAAAG

15 AAACTGAGAACCTGGTCAGTGGAGGAATTGTCTTTAAGTAATAAACGTGGGAAGGGCAGGCA

CAGTTTGAAGAACAGAGCAAGAACACTGAAATATTTGTGATGCGATTTCACTTCTATGATGTT

AATAGCACAGAGATCCCACATAAAGTGTATATAGTCAATCCTGCCTGTATCATAACTGACATT

TATATCATCAATTCAGTAACTCTATGTCACGTGACTTGAGGTTAGCATAAGTGTGAGATGATC

TTTGTCCCTACCTGATGAAACTCATGTAACTCTTTCCTGATCTGTCTGTATAACATACACATCT

20 AAATAAATGCCTAAACCTGAATTATCAGAAAGAAAAAATAGTTTTTTCAGATTCCTGATCAAA

AAATCTACGATGCACAGAATACATATAGTACCTCAACAGTGCTAGCTGGAAATCCTTTTTTGA

GGGGTCTGCAACTCTGAAGAGGATAGGGAAGAATACGATATGAAGGCTGCTTACTGCTCCAA

AAGAGTCAGACCCTAATCTTAAATGAGTCTAAGTTTGAGGGCAATTTTATCTGGGAAGCTCAG

ACTTCAACAGTGGGCACAGAATTCTGCATAAATAGGAAAAGGAAGAGGTGGGAAAGAGAGA

25 ACAAGCTAGAGGAGGAGTAGGGTCCCAGTAGAAAGGAGAAAGCTGGGTGCTATGTGAGGTG

AGGCATGGCAGCCAGGCCAGCACACGCACAGAAGTTGGAGGGTCTTCTTACCTTGTTCTTTGA

CAGAAGCTCTAGTGCCTTTCGATGAGCGCTCCACAATCAGTTGACTCGTGAAGGTCATGAATT

CCTGAACGCTAAGAAACACAAAATGTATTTATTGCCTACTTCTTATCACCTTGTCCCCAACACA

GTGGAAAGTGACCTCTGGGCTTATACATTAAGTAGACATTGCTTCTTGGTTTCATTCCTTTCCC

30 TCCCATCCCTAGTAACAAACACTCTATAAATGAGCACAAATACTGATAATTATGAATTATCAT

CACCATGAAAGCTCCATCTGTTTGCTACCTGGCTCACCAAAACAGGTGAATTTTCTGGGGGGT

TTTTCCACAGGATACAGTCAATTTTACATTTTGGTGAATGCATAATTTGGAATGCAATGGAAA

AACAAGAGGCAGGTCCTGCTCTCAAGGTCCCAATAACTTCCAAGAAGCAGGACATTTATAAG

AACTGCACTAGAAGAATAGTGTGCAAAAACTGTCAGGCAGAAATGCACAACCATTTATGGCT

35 GTGTCCACATGACAGACCCTCGCAATGCCACATACACCCATAGTGAGTGCTGGCTCAGGTCTG

CTGGGGCTCGTCCACAGAACGAGCGCAAGACACTCTGGATGGAACAAAAGGAAAACTGCTCA

TCCAAGACAAAGAAGTGGGAAATGGCTCATACAAAGGGTGAAAGGGAGAAGGTCCATCATG

GGCTCAACAGAGAGATCTATCCAGAACAGAACAGTCACAGGAGATGGTACAGCCAGAGGAA

GAGGTGCTGACAAGGAGCCTCCAACTGAGGATGTGATATAAAGGGCAACCAGGGCCATCAAA

5   GCAGGGTGCTCAAATGGGAGTCTGCAGCAGGCTCCAGCAGAGCCATATAGGTAACTGAAGGC

CTGACTCTGGGCCTGTGTGCTGTGCCTCCACATTAAAAAAATCAAGATTTGTGCAACAGTTAA

ACGAGGTAATACGTGTAAAGCACTTGGAACAATGCCTGCACACACAGTATTACTTGTTAATAT

CTTGAGGGACTGAAGTGATCAAAATAACCCCTCAGAAAAGAAGACCTCAAACAAGGAAGGCT

TTGCAGTAAACCTAGAGACAGCATTTGAGACACGGCTATAAAGAGACAAAGGAAGAACTGCA

10   TTGTGACAGCATGTATACAAAGACCAAAAAAGCTGGGAAACTACTTTTTCAACTTTGGAATCG

GGTAATTATAGGGCACAAAGGACGTAAGTAAAGCGGTCTTATAAGAAAACAAGCTCAGGCCG

GACGTGGTGGCTCAAGCCTGTAATCCTAGCACTTTGGGAGGCCAAGGCAGGCGGATCACTTG

AGCTCAGGAGTTCGAGACCAGCCTGGCTAACATGGTAAAACCCCATCTCTACTAAAAATACA

AAAATTAGCCGGGTGTGGTGGTGCGCGCCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGG

15   AGAATCACTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCTGACACTGTGCCACTGCACTCCA

GCCTGGGTGACAGAGCAAGACTCCATCTAAAATAAAATAAATAAATAAATAAATCAGCTGGG

ACATGTGTTGTTTTAAGACATATTAGTAGAGATGTCCCTTTAGTGTTGCAGCTGTTAGTCATTG

GAAACTAGTGTGGGCATCCCAAGCAGGTGAGGTATAAGTCCTACAAGTGAAATCTCTGAGAA

TCTTAAGTACTAATGGGAAGGAAAAAGGAAAAGAATCAGAGCCAAGTTGGCACCAAAAGTT

20   CCATCTGAGAAAAGCAACAACACAGAGCAGTGAATGTAGGCCATGGTAAAGACTGCAAAGAC

CAAGAACCCCAAGAAGGAGCTAAAAGATAATGCAGCAATTCCGCTTCTGGGTAAATACCAAA

AAAATGCGAGCAGGGTCTTGAAGAGATATTTGTACATCCATGTTCATAGCAGTATCATTCACA

ATGGCTGAAATGTGGAAGCAACCCAGGTGTCCACTGACAGATGAACAGATAAGCAAAATGTG

GTGAATAATACAATGGATTATTCAGCCTTAAAAAGGAAGAAATTCTGATATATGCAACAAG

25   ATGCATGAGCCTTGAGGACATTATGCTACATGAAATAAGCCAGACACACAAAAACTATATGA

TTCCATTTATCTAAGGTCGCCAGAAAAGTCAAAATCACAGAGACAAATTAGAATGGCAGTTGC

CATGGGCTGGGGGAGAAGGGAATGTGTTTAATAGACACGAATTTGATAAAAAGGAGTTCTGG

AGACGATTGACAGTGATGGCTGCACAACACTATCAATCTATTTCATATCAATGCACTCACTAC

ACGCTTAAAGATAGTGAAGATAAATTTTGTGTACCATTTTACCACAATTAAAAATATTTTTTTA

30   AAAGAACTCAAAGAAGCAGAAAGTTTCAACAAAATAACATTTTTTTTTTTTTACATCCAGCAA

GTCCTTGGCAAAGAACTCTCATCAAGAACCAGCTGCACTGAAGCAGGGAAAACAGAATCCAA

ACGGCAGATTCCATCAGATTTTGAGACAAGATGACCATAGATACCGACCATGTAGGGTCCTCC

TTCTTTCGTGCCTGAGTCACCCCAATCCCTCCCACGAATGGTCTGGAAGTGTCTGTGTTACTTC

TAACACGTTCCAGCAATTAAAGCGCCCCAGAAACAAGTAAAAGCCTGTAAGCCCTACAGATC

35   CCATGCTTCATTTGCATCTTCCGTGTGGAATCCTTTTGTACCACTAGTGTCCAACTAAAAAGCG

TTAAACCTGGCTTTCAGTTCTAGCTGGTTGTGATATAACCTCTTGGTACCTCAGTGACTTCACC

CATTAAAAACAAACAAAAAAAAGTATATCACTATCTCTCATACAGAATTGTTGGGAAGCCCC

GCAAGAAAATCAAAATATGGCTCTCAAGATGCGGCACCCAAGCTCCCAGAGTCAGAATCACT

GGGTGGGAAGTGTTGGTCTAAAATATAAATACCGAGGCCTCAATCTACTAATTCAGAACATCT

5 TGGCATGAAGCTTGGAAATCTGCACTACTTCACAGTCTCCTTAAAATTTTTACACGACAGAAA

TTTGAAAAACACTGAGTAGAGAACTATATTCTAGAATGGTATAAGCTCTTAAAGAGCTAATGT

TGGTTCCTCAAAGGTAGAGTCCACGGCCAGATTCCATTATAGGAGACCAAGCCCGGACAGCA

GACCCCGGGCCCTCCCCACCCCGCCCCGCCTCTGACTCGGACACCAGCCTTCTCAGACCCCGG

GCACTCGGCCACCCCGCCCTGCCCCTACCCTTGGCCTCCTCCACCCTCCCCTCATCCCTCCGCC

10 GACCCCAGGCCCACTCCGACTCGGACCCCCACCCCAGTCCTCTCCGCCCGACCGCCACGGCCC

ACCAGCCTGTGCCGCTCACCTGGATCTCTGGAAAAAGCTGAAGGAAGACACATCGTATGCGG

CTTTGAGCAGCACCACCTTGGCCTCGCCTTTGTAGAGGACGCTGAGGCTGTACAGCTTCATGG

TCCGGCCCTCAGGCCGCCCGCCTGCCCAGCTGCGGGACCCGTTCTCAGGGAGCAGCGCGGCCG

CCGCCCCTCGGGACCGCCGCCGCCTACCGGCCTCTCAGCAGCCGGCTGCTGACGGGGCCACCG

15 CCGGCTTCCTCCTCCTGGCTCGCAATCCACTTCCGGATCCGGTCAGCCTGGTTGAGGGTTCTCA

TACTCCGGATGCAGAAATGTGAGCCCGGAAGTACAATGCAGCGAGGGGCGGGATGCCACGCC

TCGCGTAAGCTTGGCCCCTCCCTGCTCGCCAGGTGGAGTCGGGCGCGCGGCGGGATACCGTAC

TGTCTTGTGCTGGGTGGTGCTGGGCCTCCCACAGCGGCCTGAACCCTTCTTTTTTTTTTTTTCT

TTTCTTTCTTTTTTTAAAGTAAGCATTTTTTTTATTATTATACTTTAAGTTTTAGGGTACATGTG

20 CACAACGTGCAGGTTTGTTACATATGTATACATGTGCCATGTTGGTGTGCTGCACCCATTAACT

CGTCATTTAGCATTAAGTATATCTCCTAATGCTATCCCTCCCCCCTCCCCCCACCCCACAACAG

TCCCCGGTGTGTGATGTTCGCCTTCCTGTGTCCATGTGTTCTTATTGTTCAATTCCCACCTATGA

GTGAGAACATGCGGTGTTTGGTTTTTTGTCCTTGCAATAGTTTGCTGAGAATGATGGTTTCCAG

CTTCATCCATGTCCCTACAAAGGACATGAACTCATCATTTTTTATGGCTGCATAGTATTCCATG

25 GTGTATATGTGCCACATTTTAGGAGGAGCTTGTACCATTCCTTCTGAAACTATTCCAATCAAAA

GAAAAGAGAGAATCCTCCCTAACTCATTTTATGAGGCCAGCATCATCCTGATACCAAAGGGT

GGCAGAGAGAGACACAACAAAAAAGAATTTTAGACCAATATCCTTGATGAACATTGAAGCA

AAAATCCTCAGTAAAATACTGGCAAACCGAATCCAGCAACACATCAAAAAGCTTATCCACCA

TGATCAAGTGGGCTTCATCCCTGGGATGCAAGGCTGGTTCAACATACGAAAATCAGTAAACGT

30 AATCCAGCATATAAACAGAACCAAAGACAAAACCACATGATTATCTCAATAGATGCAGAAA

AGGCCTTTGACAAAATTCAACAACCCTCATGCTAAAAACTCTCAATAAATTAGGTATTGATGG

GACGTATCTCAAAATAATAAGAGCTATCTATGACAAACCCACAGCCAATATCATACTGAATGG

ACAAAAACTGGAAGCATTCCCTTTGAAAACTGGCACAAGACTGGGATGCCCTCTCTCACCACT

CCTTTTCAACATAGTGTTGGAAGTTCTGGCCAGGGCAATCAGGTAGGAGAAGGAAATAAAGG

35 GTATTCAATTAAGAAAAGAGGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATC

TAGAAAACCCCATCGTCTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCT
CAGGATACAAAATCAATGTGCAAAAATCACAAGCAGTCTTATACACCAATAACAGACAGAGA
GCCAAATCATGAGTGAACTCCCATTCACAATTGCTTCAAAGAGAATAAAATACCTAGGAATCC
AACTTACAAGGGATGTGAAGGACCTCTTCAAGGAGAACTACAAACGACTGCTCAATGAAATA

AAAGAGGATACAAACAAATGGAAGAACATTCCATGCTCATGGGTAGGAAGAATCAGTATCGT
GAAAATGGCCATACTGCCCAAGGTAATTTATAGATTCAATGCCATCCCTATCAAGCTACCAAT
GACTTTCTTCACAGAATTGGAAAAAACTAAAGTTCATATGGAACCAAAAAAGAGCCCGCATT
GCCAAGTCAATCCTAAGCCAAAAGAACAAAGCTGGAGGCATCACACTACCTGACTTCTAACT
ATACTACAAGGCTACAGTAACCAAAACAGCATGCTACTGGTACCAAAACAGAGATATAGAGC

AATGGAACAGAACAGAGCCCTCAGAAATAATGCCGCATATCTACAAGCATCTGATCTTTGAC
AAACCTGACAAAAACAAGCAATGGGGAAAGGATTCCCTATTTAATAAATGGTGCTGGGAAAA
CTGGCTAGCCATATGTAGAAAGCTGAAACTGGATCCCTTCCTTACACCTTATACAAAAATTAA
TTCAAGATGGATTAAAGACTTACATGTTAGACCTAAAACCATAAAAACCCTAGAAGAAAACC
TAGGCAATACCATTCAGGACATAGGCATGGGCAAGGACTTCATGTCTAAAACACCAAAGCA

ATGGCAACAAAAGCCAAAATTGACAAATGGGATCTAATTAAACTAAAGAGCTTCTGCACAGC
AAAAGAAACTACCATCAGAGTGAACAGGCAACCTACAGAATGGGAGAAAATTTTTGCAACCT
ACTCATCTGACAAAGGGCTAATATCCAGAATCTACAATGAACTCAAACAAATTTACAAGAAA
AAAACAAACAACCCCATCAACAAATGGGCGAAGGATATGAACAGACACTTCTCAAAAGAAG
ACATTTATGTAGCCAAAAAACACATGAAAAAATGCTCATCATCACTGGCCATCAGAGAAATG

CAAATCAAAACCACAATGAGATACCATCTCACACCAGTTAGAATGGTGATCATTAAAAAGTC
AGGAAACAACAGGTGCTGGAGAGGATGTGGAGAAATAGGAACACTTTTACACTGTTCGTGGG
ACTGTAAACTAGTTCAACCATTGTGGAAGTCAGTGTGGCGATTCCTCAGGGATCTAGAACTGG
AAATACCATTTGACCCAGCCATCCCATTACTAGGTATATACCCAAAGGATTATAAATCATGCT
GCTATAAGGACACATGCACACGTATGTTTATTGTGGCACTGTTCACAATAGCAAAGACTTGGA

ACCAACCCAAATGAACCCTTCTTTTTGCTTGCGTTGTTGAAAGAAGGCAAGTCTATGGATAGG
AATGAGTGAGGCACAGCTCCCTGAGGATGCCATATCTTGCCCGTTTCTTGTGTATTAAGTGAC
ATCACGTGTTACCAAACTAAACCGGCTGCATTTGCCTGCGCACAACATAAAACCAAACACCCA
AGCATTGGATTTTTGTAGCAAGAAAGATGTATTGCCAAGCAGCCTTGCAAGGGGACAGAAGA
CGGGCTCAAATCTGTCTCCCAATACTTGCTTCGCAGCAGTAGATTTAAGGGAGAGATTTTGGA

AGTGGAGTTTCGGGCTGGACGGTGATTGGCTGAAACGAAGAAGTGTTTAGAAAATCTCTTGGT
CATGAGCTGTTGCTTCTTCATGCTGCTTCAAGGGTCACATGCAGATTCAGGAGGTGGTATAAA
ACAAGCTGTGGGAATTTGGGCTGTGACATCAAAGGGCCGCTCCTCGGGCTAGTAAGTCTATTT
TGCACAGGCTCCAGTCAGCCATATTGGTTCCAACCTGTTCCAGCAAGTTGTATAAGCAGAGGG
GATTATAGCAAACTGTTTCCTTATCGGCTGCCCTGCAAGACAAGCTCAAGATTTCTGTTAGTTA

CCAGTTTCTTTAACCCTGTCGGGCACAGTTTCACATGTAATCAGAAAGGAACTTGCAAGACAC

ATACAACTGAAAGAAACTTGGTCTTTGGAAGTTGTCAGTAAGGTCACAAAGTTGTGATGCTAG

AAGCAGCCGTATCTGAGATTATGGGAAAGAGATGATATATTGGAAAAACAACAGCATCACTT

TAAACATTACTCTAAATCAAGGTTTCTCAACCTTGGCACTATTGACATTTTGGGTTAGATAGTT

CTTTCTTGTTGGGAGACTGCCCTGTACATTGTGTAGGCAGCATCTCAGGCCTTTGTAGAAATGT

5   CAGTACCAACCCACCCCCTCCCCACTGCACAATCAAAAACGTCAAAATGTCCTTTGGGAGCAG

TAGTTTTGAGAAACATTGCTTTGCAGATATATATGTTTGTTTGTTTGTTTTGCTTTGTGACAGG

GTCTTACTCTGTTGCCCAGGCAGAAGTGCAATGGTGTGATCCCACTCACTGCAACCTCTGCCTC

CCAGGTTCAAGCGATTCTCATGCCTCAGCCTCCCGAGTAGCTGGGATTACAGGAATGCATCCA

TACACGCGGCTAATTTTTGTATTTTTAATAGAGATGGGATTTCACCATGTTGGCCAGGCTGGTC

10 TGAAACTCCTGGCCTCATGTGATCCACCCACCTCGACCTCCCAAATTGCTGGGATTACAAGCT

TAAGCCACTGCGCCCAGCTGAGAAACATTGCTTTAAATAATCTGTGGTGAAAGGAAGTTCCCA

CCACCTGCCCACTCACTCAGTACCTCTGTCACCAACCCTCTTCCCTGGGTGTTTCCAAGTACAG

AGGGTGGAAAGGGCTTTTCCACATTTCCCCTGTTTTGGTAGTAAACATTAGGAACAGCCATTG

GCCGTGGCTAGGCTCAGCCACCCACAGATATGGACACAGTAGTCTGACAAGCTGGGTTGCTG

15 GGTGCTATCAGTCCAGGCTCAACTGCTTGCACTGACACCATTTCCCTATAGGAGGCAGGTGAG

AGCCATTTCTGAGGAAAGTCTCTGGAGCCCCTCTTCCTTCCACTGAAAGTTGTGCAAAAAGAT

CAGGAAGACAGCGCTTGGATGGAATAAATTTCAGTGTATCCACTTGACACATTATAGTGGCTG

TCCCAAAGTTTACCTTATGCCAAGTACTTTCCATGTGCCACATCATTTAATCCTCACAAAAACA

GGGGAAAATATTATTGCCACCCTACAGACATAGAGACTGAGATTCAATTTAAGGAGATGGTT

20 GGTAAGGGACAGAGTTGGGGTTCAGATGTCAACAGTGAAATGCTTAACAAACTGTCATGCAG

CCCACTCCTGGCAACTCTTCCTGCTCCTCTCTGGCCTCACTCAGCCTCTACTGTTCCAGGAAGC

CTCATTCATAGTCATGTGGTTGCAGACTTCCCAAGCTCACTGTGTTACCAAAAAGCAAGACCT

GCCTTCTGCTGCATCGCCCCAGCTGTCACCCAACTTGGATTCAGTCCCAGCACTGACACATCA

CAAAATCACAAAGTGAGCAAACCATTACCTCCCTGAGTCTCCTTTTGTTTTTATCTATAAAC

25 TAGAAAAATATTCTTTCCATAGGAATGTTGTTGGAAATAATAAACATTATATTACAAGCTCT

AGTCATTGTTGATGTTTAACAGGTAACAGTGATAATTATTTGTCTTCTCATTAATGAAGAAAA

GGATTATTAATCATAGAGGGTGGAAGGCATCTATGGGAAGTAGAGATTTGAAGATAGGCTAA

AACCCAAGTAAGGCCTCTAGATTAGATAATAGTATTGTATCTATTTTAATTTCCTGCTTTCCAT

CACTGTGCCATGGTTATATAAGAGAAGTCTTTGTTTATAGGAAATATACACAAGAATTTAGAA

30 GTAAAGGGACATTGTGTCTGCAACTTACTCTTACAGGGTGTGTGTGTGTGTGTGTGTGTGTGTG

TGTGAGAGAGAGAGAGACAGAGAGAGAGAGACAGAGAGAAAGAGAATGATAAAGCA

AATACAGGAATCAGGATGAAGCGTATCTGTTTGTTTGTTTTGCTTTGTGATAGGGTCTTGCTCT

GTTGCCCAGGCAGGAGTGCAATGGTGTGATCCCGCTCACTGCAACCTCTGCCTCCCAGGTTCA

AGCGATTCTCATGCTTGTATTGTTCTTGCACCTGTTCTGCAAGTACAACATTGTGGGAATGGAA

35 AATGCAGGAAATGGGCAGTAAGGCTATGAACGAAGCCCGCACAGGAGTGTGGGTAGCAGAG

TTCTCTAGTCCAGGCTCCCACCTGAGGTGCTGGGACCTAGAAGAAAAGCCTCTCTGCAGACAG

AACTGGAGTTAACGCTGTCCACGATAAATGGCCCAGGCCCTGTTAAGTTTGCCCCATTGAGCA

AAACAAGTACCCACCCGCCTTTGCAGCCTTGCCTAGCTCACATAAGGTGCCAGCCCTTGCTGT

ACAGCAGAACCTTTGGGGAGCTGGACAAAAGCCTATCAAGGAGCATACCCCCAGGAAGCCCA

5 GTCCAGGTGGGGAGCCCAGCCACACAATGGCCCTTGCCCCCACACCTCCTCATTCAGTCAGCT

AAGGCCATGGCAGCTGAGCTGCCTCCACAGCTCATATAGGAAAAGGGTGTGGAAAGGGGCCA

CCAATGTGGTCAGGCCTCCATGGCCTGAGTAGGTCACCAAGCCTCAGGTGCACAGACTTGATG

TCATCAATCAGGGTCTGTCAGCACACCTAGCCCTCAGGAACACTGCTCCCCACTGCAACCCCA

CACCAAGGCATCCTGGGCTCCCTCTGGGTTCTCCAGGCCCCAGGGAAGACAGACAGAGTCTGC

10 CACCAAAGGTTTGAGCTCTGCCACTGGCTACGAAGCAATAGGGGATGTCAGAGCAAGGGAGG

AACAGGACAGGAGTATACGTGGGCAGGAAGGGATTACAGCCAAGGAAGACAGGAGGCAGGT

GCCCTGATTTTGAGGCTGTGCCCCAGCAGGGGCTTCCCAGAAGCTGTATTTGTCCTAAGACAC

CCCTCTGCAGCTGAGGGGCTAGAGATGGATATGTAGCTGTGTTAGGCCATTCTTGCATTGCTA

TAAAGAAATACCTGAGACCAGGTAATTTATAAAGAAAAGAGGTTTCATTGGTTCACAGTTCTG

15 CTGGCTTTGCAAGAGGCATGGTGCTGGCATCTGCTCAGCCTTTGAGGAGGCCTCAGGAAACTT

ACAGTCATGGCGGAAGGCAAAGGGGAAGCAGGCACATCACACAGTGGAAGCAGGAGTGAGA

GAGAGAGAGGCACTGGGAGGTGCCACACTTTTAAACAACCAGATCTCGTGTGAACTCAGAGC

AAGAGCTGACTCATCACCAAGGGGATGGCCCAAGCCATTCATGAGGGATCCACCCCCATGAC

TCAGACACCTCCCACCAGGCCCCACCTCCAATATTGGGGATTACAATTCAGATGAGATTTGGT

20 GGGGACACATATCCAAACCATATCAGTTATCAGTAGCCATACTGGATGAATGCCAGGAACTTA

GAATTAGGACACATGGTCATTTAGGCAAGTGGCTTGTCCTGTCAATGGTACCCTGATAGTCGT

GGGGTTGCCCCGTACAAAAAGCGAGAGGAAGTCTACAGAGCTGTCAAAGAGGGGCAGGTGG

AAAGGCCTGCAGAGGAGTCCCCTGCTCCACAACCAGGCGTGCACCTCCCACATCCTCGGGGCT

GTAGGCCCCACATGAGAGCAGAAAGAAGGATGCAGAGGAAGGCC

25 AAGAACACAAGGTGTGCCCTTGGAAAGGCTGGGCACACCAAACACAACCTAATAAACAACAG

CAATGAGCACACAGGGAAAGTACTCACAGGGAAACCATCATGAACTAGAGGCTGATCCCACA

CCCTGCCACATGGGGCCCCAGGCCCCAGCCTATCAACCAGTGGTCCTTATTGCCACAGCGATT

GGTCTTTGGATAGGCACCTGATGCAAGCTTCAGCCAATCAACAGGCCACTCAGCTGGCCATCA

GTAGGCCATCCAATCAGAGCAAAGCCCAGGACTTTCTTCGACTCTTAAGAAAAGAGAAGCAA

30 AGTAACTGGCACAGATTGGAGAGGATCAAGGAACCCCGAGCTGGATACATACAAACTTTGGG

TTAACATGGATGATTAAATACATATGTTTATGTGAACCACCTCCCAAATATGCTCCACTATAAT

GACACAAGACAAAGGGCAGGGGGAGACCAATTGCAAGGTGGCGCAAATGAGAGATGCTACC

AAGGGTGGCGGGGGAGAGAGGGGAGCAGTTGTCAAGTTAGGAGGCAACAGGCTGAGGGACA

GGGACCAGCAGACGGGGAGGGAGGGGCTGAAGCAGAAGTGTCCAGTGTCTGGAGGGATGGG

35 GCCAGAAAGGCAAGGGGCATCCTGAAGAAGCTATACCTGGGGAGGGCAGCTCTCTCCCCACC

15

TGCTCCCCAATTCATCAGCCAGGAATGCCCCATCCACCCCACCCCAGGGAGGAGGACAGAGG

ACTTTCGTTTGGGAGCATTGAATGGTTCAGAGATTCTGCAACTCTGCGGTCCCCAACTAAACT

GCTCATTGTTTCAAGCAGTCCCTGTTGGGTAAATGTCCCCCATTGTAACCGGACTCGGATTCCA

CCGCTTGAAAGCCAAATACAAGAGGAGAGGTTTGGTGGGAGGAAAAGTGGTTTTAACTAGAG

5   CCAGCAAACCAAGAAGATGGTGAATTGTTGTTTTAAAGCATTCAATTATCTCAAATTTTAAAA

TTTATCATAGGATTCTGAAAGGAAAACTTGGTATGGGACATACGTGGGAGCAGTGCAGGGTA

CAGGGTCTATGTGTCTTGATCCAATGGCTGTCTTGAGTATCACCTATCCTGAGGTCTGGTTGGT

GTTATCTTTCCTTCGGCCAGATGGTGGTGGGTGAATTGTTTCGACTCCCCCTAAGTTGGAGGAT

TCCGCAGGGGTTCCGTGTCTGGTTTTTGTTTCAAGATTAGCCCCTGGAATTCCCAAATAAGCAT

10 AGAGTTAGATAAGCGGGCATGGTGCAAAGGAGTGTCTAGTGGGAAAGGGAGAGAAGCAGAG

TTTCAAAGTACATTTCAAGGTTACATTTTAAGACTAAAGAAAAAGCCTTAAAATGCATTTTTA

AAGCTGATTTAATGCTTGGCTACACTAGGCTGTGGCCAGTGTGCAGTGTGGCTGCTCTTGGAT

CAGGTGATGTTTCATCAGCTGTGTCCAGGGAGGGCAGGGCCATGTGGCAGAACCTGGGACCT

CTGTGTGAGGGACTACCTTGGCCCCTGTCCTTAGCAGGAAGCTATGGTAAGGAACCCTTAGGG

15 AGACATTAAATTGGGGAGACCGTCCCTGCCAATCCTTTAACCTCCCCAGCCTCAGCGACCTCA

GTTGGAAAGTGGTGGTAATAATACTACCACTGACCAGGTGTGGTGGCCAGACATTCCACACTT

TGGCTTCAGCCGCTCCCTCCCCACTCTACTGTAATCCCAGCACTTTGGGAGGAAGAGGTAGGC

GGAACCTGAGGTCTGGAGTTGAGACCAGCCTGGTCAACATGGTGAAACCCCATATCTACTAA

AAAGAAAGTACAAAAAATTAGCCAGGTGCAGTGGCACACGTGTGTGGTCCCAGCTACTCGTG

20 GGTCTGAGGCATGAAAATTGTTTGAGCCTGGGAGGCAGAGGTTCCATTGAGTGGAGATCGAG

CCACTGCACTCCAGCCTGGGTGATAGAACGAGATTCTGTCTCAAAAAAATAAAAATAAAATA

ATAATAATAATACCACTGCCTGCCACACTAAGATTGTCTGATTAGATGACAGAATGAATGCAA

AAGTACTTTGTGAATCATAAATGTTTTCATCAATATTAGTTATAATGACAATTGCTCCTTCTCC

TAATAAATGTATTGCCTTTCTTTAGGAATAAATATAACAAGAAATGTGTAAGATATATATGAG

25 AAAAATAATAAAATTCACCTGAAGGACATAAAAGAAGACCAAAATAAATGAAACAACACAT

ACTTCTAGATGAGAAAACTCAATATTATAAAGAGGTTAGTTCTCTAAAATGAATCCCTAAACC

CACAAAGTCAATGTATTTCCAATGAAATTGTCAACAGCATTATTTTCCGAAGTGGGATGAGTA

GTGCTAAGATTTATAAGAAAGCCAACATTCCAGAGCAGTGGGGAAGGGATTGCTTCACCACC

AAATAGCCATATTAGAGATTCCCTTGCACCATACCCAAACCACCATCTCCCAGGACCCGGGAG

30 AGCAGAAAGAGGAATGAGAAGAAAGGCGAGGATGTGAGGTGTGCCCTCATAATGGCGGTG

CACGCAGCACAAGCAATTGCAGAAGACTAAAGTACTGAACAAATAGAAAACTTGGAAAAAT

ATTAGAAGGAAATGTGGGAGAACATTTTTGCAATTTGGGGATTGGAAACGGTTTTCTTAACAA

GATATAAAAACCCCAAAACAAGAAAACAAAGGTTGAAATTCATAAAAACTAGATACTTCTGT

ATGATGAAAGACACGATTAATCAAGTTGTTAAGTTTAGCAATAGACTAGGGGAGATATCATA

35 GTATATTTAACAGACAAAGGATTAATAGATACTACAGATGAAATATAAAATAGTTTCTCCAAG

16

TCCATAGGCAGAAGATAATCCAATAGCAACATAGTTAAGTAATGTAAACAAATCATCCTTAG

AAGAAGAAATGCAATCACCAAGAAACACATGAAAAGGTGTCCAGCATTTTGCAATTCAAGCA

ACAATGAGGTGACAGATCGGCAAAAAACTCATAAAGATTTATCATCTGAAGGATTGGCCAAG

ATAAAGCCAAACTTCTCGTGTTGGCAGAAGAAACTGGTGAAGCCATGTGAAGAGGCCACGTG

5   GTCCTGCCTACCAAGATGTAAAATGTGTACAGCATTTGAACTAGCAATTCAGCCTCCAGGAGC

CATCCAGAAGAAACACTGACACACTTAGACTCCGGTGAAATTCAAGGACTTCTGCCACAG

CCTGCTTCGTAATAGTGAAAATCTGAAACTGCCTCAATGACCGTCAATAGGAAGTTGATTTTA

AAGTGTTACAGCACATCTGTCTGGAGAGATCGCACTGGCCACTCCTCCTCACCCCCTCTGCTG

GACCTCTGAGCGTAGGTGGCCTGGAGCTGGGTCCTGAGCCCTCTTTGGTCTATACCGACACTA

10 CCCAATATGGTAGCCACCAGTCACGCTGGACACTTGAAAAGTGGCCGATCCTGACTGAGAAG

GGCCACGAGTGGGAAAAACACACCAGACCTCAGTGACTTAGGCAGAAGTATGTTTTGTTCCA

GACTATTGACTGAGCCCGCAGCTGAGTTGGCTCCAGCACCCTGGCCCCCTGCTCCATCCACTC

ACTGGGACTCCCCACTGCACAGGGCAACCTCTCCAGGGGCACTTGGGCTGCGAAGGGGAGAG

TGGGTGGCATCCCAGGCTGAAGCTTCCTGAGCAGGGCCAGAGGAGGAGCCAGTCCCTGTGGG

15 CCTCTGTTCTGACAGTGTCAACCTCAGCCAGGCTTGTGTGGGCCAGGTGTACTGTTCTGGTTCA

GATTTCAAGGAGATAGTCAGGGCAGGCCGCGCCAAAGCCCTCCGATGGGCTCCCCTACTGCCT

GGCAGACCTGTCCAGCTTTGGACTCTGGCCCTGCGACCTGGAAGTCAGGCTGCCAAGAGGTCC

AGGCAGTGGCCTCCACTGTGGAGGGTCTCTGGAGAGTTTACAGCCCTAGATAGGGGGGTTAG

GGATGTGAGATGGTCCCAGGGGCCTGCTCCTGAGCCACGCCAAGCTGCCTGCTCCCTTTCCTC

20 TGCTTCCAGACTCACGGGATCCTCTGCTCATCAGAACAGGAGTGTGGGAGACCCTGAGACACT

GCCCCAGGATCTGAACAGGTGGCAAAGGCTTAACAGGCTAGCGGTCACTGTAGTGACAAGGC

GATTGAGTGGTCACCATGGTGATGGGGATGGAGGCTCTTTGCCACCAGTCCCAGTTTTATGCA

TGGCAGCTCTAATGACAGGATGGTCAGCCCTGCTGAGGCCACTCCTGGTCACCATGACAACCA

CAGGCCCTCTCAGGAGCACAGTAAGCCCTGGCAGGAGAATCCCCCACTCCACACCTGGCTGG

25 AGCAGGAAATGCCGAGCGGCGCCTGAGCCCCAGGGAAGCAGGCTAGGATGTGAGAGACACA

GTCACCTGCAGCCTAATTACTCAAAAGCTGTCCCCAGGTCACAGAAGGGAGAGGACATTTCCC

ACTGAATCTGTCTGAAGGACACTAAGCCCCACAGCTCAACACAACCAGGAGAGAAAGCGCTG

AGGACGCCACCCAAGCGCCCAGCAATGGCCCTGCCTGGAGAACATCCAGGCTCAGTGAGGAA

GGGTCCAGAAGGGAATGCTTGCCGACTCGTTGGAGAACAATGAAAAGGAGGAAACTGTGACT

30 GAACCTCAAACCCCAAACCAGCCCGAGGAGAACCACATTCTCCCAGGGACCCAGGGCGGGCC

GTGACCCCTGCGGCGGAGAAGCCTTGGATATTTCCACTTCAGAAGCCTACTGGGGAAGGCTGA

GGGGTCCCAGCTCCCCACGCTGGCTGCTGTGCAGATGCTGGACGACAGAGCCAGGATGGAGG

CCGCCAAGAAGGAGAAGGTATCTCGCCCTCCATTGGGCATTCTGGGAGTGTTTGCTTGCCTGT

CCCCAACATTCCATGGTTTGTTTGAGCCTCAGAATCTGATTTTATGCACAGGCTCTTTGAGAAG

35 GGTCTTGCCAGGGGTGCCTTCTGGGGCAGGAAGGCCCCTACTGCCTGGCAGACCCATCCAGCT

17

TTGGACTCTGGTCCTGCGACCCGGAAGTCAGGCTGCCAAGAGGTCCAGGCAGTGGCCTCCACT

GGGGAGGGGCTCTGGAGAGTTTAGAGCCCTAGATGTGGGGGGTTAGGGACATGAGGTCTTGTG

GACAAAGCCCACTACCTGATTTTGAGACAACACTCACTAGACATGGTGACAAGTCAAAGATG

CCTTGCCTCCTACCAGGAATCACTTCGCAGGGAGCCCGAGGGCTGCTGTGGCCTGCTGAGGAG

5   TGCAGGGCAGTTACTTTTTCCAAAAACAAAGAGAAATCCAGGCATGCTCTGAGCCAGCCCTGA

GCCCAGCAGTGAGCAAGGAGAGAGCTGGAGACAGGGGACTTTGCTGTGAAACACTGGGGGG

AATGTGCCTGCATCACCCCAGCTGGGGGCCCAGGCAGAGTGGGGGAGAAGGGGTAAGTGGGC

AGAGCCAGTCACTTTGGGCATGCTTCCCTCTCGCCTCTGTGTGAAATGACCAGGTCAGCATAA

ACCCCGGGCTGGCTGTGCTTCTGGCAGAGCTAATGATGTTAGGAGGAAAACAACCAACCCAA

10   GTGAGAGGGTGCGCAGCCAGACAGCTGGACCGGCCGAGGCCCCAACCAAGTCCCAGATCTGC

CTGTCACTGGTGCTATGGCAGCAATTTGGATGAGAAATCCTGCCCAAAGGGCCCCTTCAGGCC

ACCCGGGGAGAAGGAAGCGGCTGTCTTTGGCATGACCAGAAAGATGGCTCGGAGCTAGGGAG

AGGTGGACATGTGGGCTGTGGAGATCTGGCACTTTCCCCAAACAAGGAGAGAAAGCATAGTG

TGCCTATGTGTGAATGTGCTATGTGTGCATGTTTGTGCCTGTGCATACCTGCATGTGTACATGC

15   ATGTGCACATATGTGTGCACAGGGAATCACTTTAATAAAGGCCACAGCAGAGCTGTCCCTGAG

CCCCTTGCATTCACAGTGGCATGTGAGTGAACCACCTTCTTAGGCTGGGCATCCAGTCTCAGA

CTCTGGGGCTGCCCATGCCCCATCCTTTATCTGCTCCACGTGTGAGGGGTTGCTGGTCCTGACC

AGGGCCAGCTGTGAACCCCAGAATCCTGGGAAGTCACTGACATTCTTGTCAGGGCCAAGAGT

GGAGCAAGGCAATGCCTCGGGCACAAACTTTAAGGGGTCACCAGAAACATCAATCATCAAGA

20   TATATGCTATTTTAAATAATCAAAATGAATGCAAAAAAAATTTATGATGGACAACATACCAAA

TTCTAAACAAAGGCAGGATGAGTATCACTGGCTTCTGCACTTTTCTCCACCCAGTCTACCCCTC

TTCTAGTGCCTGGATCGCAGGGTGCCAAGGCCTGGATGAGGGAAGCGTGGAGCTGCAATGGC

CACTCCTGTCTGCCTGTTCTGGCTGCACAGAGGACTCAGTCCTTGTCTTGGGGGAACCTATCTT

GGTTTTAGGGTCATCCTAAGGATCTGATGTTTTCCAAGTGAGCTGGCTGTCCAGGCCCACCCA

25   GGTTCAGTCCAGTCCTGTGTCTCTGGGAAGTGCTGCCCCTACCCCAAGCCAGTGTTTGACCTTG

GAGCAATGAGCAATGCCCTCCTTCCACTTTCAAAGTTGTCCCCAAGACGTCAGCTGTGGTTGT

CTCTGTGCAGACACCGAGGAGGAACTGTCTTCTTTCTCCTTTTGGTTGCTTTGGAGGAAAGTAA

AGTGTTGCTGGTTTCCCTCTTTCTACTTCTTTGATTGAGAGCAGCCGTCTTGCCGGTACCAACC

TTCCAGATCTTACCTGTGGTTGCAGGAGCCTGTGGCCTCAGTCCTGTGCCCAGTGACTTCTCCA

30   TGTGGATGTCAGCTCCTTAGGGGCAAGCCTGATTCCACTGACACTACTCCCACCCCTCATAAG

CCCCTTCTTACCAGCTGCAGTTGCCTGGTACCCCACCATCGCTGACTCATTCCTTTGGCATCAA

GGTTCATCCCTTACTGGGCCACCACTTCTGGGTGGCCTGAAATAGGGCCCTGGGCATCCCTCTT

GGGGACCTTTTGGTCTATATTTTCACTCTCACCTCACTAAGGACAGATGAGTAAATCTGGTTAA

CTTTGCCTGATAGATTTGGTGACCTTTTTTCAGGAAGGAGCCTGGAAAGATGAGATTCAGGTG

35   TATTGGTCAGCTTAGACTGCCATAAGAGAATACCATCCACTGATGGCTTAGAAACAACAGAA

18

ATCTATTTCTCACTATTCTAGAGGCTGGACGTCCAAGATCAGATGCCAGCATGGTCAGGTTGC

AGGGAGGGCTCTCTTCCTGACTTGCAGACCGCCACCTTCTTGCTGTGTCCTCACATCGTGGAG

AGAGAGTGAAAACAAGCTCTCTGGTGTCTCTTCTTATAAGAATGCTAATCCTATGATGGGGGC

TCCCCCTCCTTACCTCATCTAAACCTAATTATCTCCCAAAGGTCTCATCTCCAGATACCATCAC

5   ACTGGGGTTAGGGCTTTGACATATGAATCTGGGGGGACACAATTCAATCTGTAACACCAGGA

GGGCATGCCGGGAGGAACTGACCTTCCTCCCTCCAGCTGCCCTGGACACCTTTGCCCCATTGA

AGGAGCAGGCTCAGAAGTGGAATGAGGATGGAATAAGGTGCACTCCATCATGCTTACCCACA

TCCCTGGCAGGAATTGTCCTGGGCCCCAGCAGGAGAGATGCCCCCCCATACTGCCATGGCACC

TGCTCTGAGACAGGTGTGCAGAGTGCAAAGCTCCAGGTGGCCCCCAAGCAGGTGTGCTGGGA

10 GGAGGGGCCCGTGTGGGAGGAGCAGGCAGCGCCAAGGCCTAGCGGAGCAGTGACAGGTCCC

TGACTTCAGGGAATGGGCACGCTGTGGGCAGGCAGCTGGTGTGGGGGTGAGGGCTGGGGCTG

CATCTGTGGGACCAGGGCTGGGCCATCCTCATATGCCGTGTCACAACCCCAGTGCCCCTGCTG

TAGCCAGGACAGGAGGCTGGGCCAGGCTGGGAGGTGACAAGAGTGGGGGCTGTCCCCAGGA

GAAGCACTCTGCTGCCTGTGCCCAGGCCTCTGGGGATGAGGACCCCTCAGAAGGAGTAGCTAT

15 GTCTAGGAAGCCCCAGGGCAGGAGCAAGCCAAAGGGGACATCATTAGTGAGATCCAGGGGAT

CAGTGGGCCACAGAAGCCCCAGCGTGAGCCCCTCTGACTGATGCAGCTAGGCCCACACCTGC

ACCTGCCCACAGCAAGACCCCCAGGAGGAGAGGGGACAGATGGAGAGAGGCACAAAGTGCC

CCTGGCCTCTGCCTTGAAGCCACCCCAAGGCAAGAGAGATTTGAGCCCCTGTTTAGTGACCTC

CAGGGGAACATTCTGGCCCATCTGATGTGGGAAGCCCCTTGTGGAGTCTGTCATTCCTCAGCT

20 GAGCCAGGCCTTTGGAGGCAGCCCAGGCATGTCCCCTGTGTGCTCCTATCCCTGTGTTGGGAC

ACCTGGCCCAGCCCCTCCTTCTGCCTTTCTCTTCCCTTCCCTTCTCAGGAGTGGACACTTCCTCC

TTTAGCCCCCTCACAGCTGTGTGAACTTCTCTGTATCTCTCTCTTTCTGTCTCTTTCTCCCCCTCT

CTCTCTGTCTCATTGTCTCTCTGTGTGTCTGTCTGTAGTATTCTCTCTCTGTCTCTGTCACTCTGT

CTCTCTCTCTCTGTGTCTACCTTTCTGTATTTCGCTTTGTTTCTTTTTCTCTGTGTGTGTGTGT

25 GTGTATCTGTTTTTCTCACTCTCTCTCTGTGTCTATCTTTCTGTATTTCGCTTTGTTTCTTTTTCT

GTGTGTGTGTGTGTATATCTGTTTTTCTCACTCTCTCAATCTCTCTCTCTCTTTCTGTCTCTCT

TTTGCTGGCCTGAGCAAAGAGGGAGCCCCATCCTGATGCTACATAACCG

TGAACCAGCACAGACAGAATTGTAGGAAAGTCCTGCAAGTAGAAGGATAGAAGGATGAGGG

AAGAAACGCCATGTGAGTCATGACAGATCCCTTTCCAGGAGCCACTGACTCACCCTGCCTCCT

30 GCCCTCCCACTGTGACACTATTACTCACAGACAGGCCCGGATTAAACCTATGTTCCAGGTGCC

CTGTGGTTCCCACAGTGTGGCTCCCTGGGTCTGGCCTCAGGCTCCACAGGTGCCCAGCCCTGC

CAAAGTCTCCAGAGCAGCTGTCCAGCTGGGGAGCTGCGGGGCCCCTTCACAGAGCGCATGGG

AAGAAGTTCCATCCTACACATTACATCGAGAGGGACGTGCCTGAGAAGGGGAGCTGGAGCCC

GTGCAGCCCCTGCTTGCGTGCAGAACATAGTGTACCCTGAGCATGCCATGAAAAACACAAA

35 CGCACAAAGTTGTAAAGAAAAAGAAATGACAGGTGGCTGTAAAATCAGTTATAGCCCACGA

GAGGCCCACTAATGAGTGGTGATTTCAGCTGATTACAAAGAAATGATGGTGTTTCTGTAATGA

ACTAAACATGCACTCGTGCGTGCACACACGCGCACGTATAGTCACATAACTGACCAGCCCTAT

GCATCACTTGTTAATTACTTAGTAACTGTAACAATAATAGTTTCCAATAAGTGAGCCTTAGTCT

CTGCGCAAGGGTCAGTTTATTGAGCACACGGGGGCCTTGCAGTGGGGGCAGGTGATCTGCTCC

5   TGGGAGCCGCCAGCCTCTCCTCTCCTGCTCTTCATCTTCCTCCGTGGTGGGAAATTGTCTCACT

GCTTCTACACCTGAGGCTGAACATCTCCCTTTATTTCAGTCTGAAACACATGTAAAAATATACT

GGAATGAATTAAGGTTGCAATTATTGATATCAGGCAGTGAGTACATCAGGGTTTATTATACTA

TCTCCTTTACTTACTTCGAAGTTCTCTATTACCAAAAAATTAAAAACTATAAAGAAAGAAAA

AGGAAATGAGGCTAGATTCAACACAGATTACTCTTACCAAACCCTTCGTAGTCCCAGGAGTCC

10   CCTAACACAAGCACTTGTGACCTGGAGTGATATTCACAGCATTCCTTACCTGGCAATACCTGA

GTATTAGCCCCCCCAGTGGGGATCTTTGTTGTAGACAACCAGCAACTATCAGCCCAGCCAATAA

ACAAGTAGGAAAGGGGAGTGCTGGAGAGGCCAAGAAGTGGGATTTTCCATGCTCCTGGGCTG

TGATCCAGAGGGCACGGCTGTGAGGCTGATCTCAATGAACACTCTGTCTTGGAAGTACAGGG

ATCCTCTGCTACCTGAAAACGTTCTGAGTATTCACTTTCATGGATTGCAAAGTCATTTACCCAA

15   AATTCACTCTCCAAATGAAAAGTGAGTATGATGAATCAGTATTCAAGTTCCACCTGGGTCCTG

GGAGAGGGCATGGACATCATATCCCAGCTGTTCCGACAGGAGGACCCAATCTGAGTCTCACT

GCCTGCCTGCATCGTTTGTCTGCTGCCAGCCTGCACAGTAGGAAGGGAAAACATGATTTGTAT

CTGTTTTAGGTCAGGTTCCCAAGAAGTAGAGCCTGAGATTGGAATTCTTGGAAAATGGTGTTT

GCGGGAGCGCTGTCAGCAGAAGCTATAAGGAAGTTGGGGGGACAGAAACGAGAGGTAAGA

20   AGCCAGTCAAAAAGGCAGGTCCAGCTTAAGTCCGCCTCAGTCTGGTTCCACAAGGGCTCTGAT

GCATGAAGAATATCACAGGGTTGTCCCTCCTGGGAGAGGGGCCAGCCTATTGTACCTGTATCA

AAGCCACCAGCTGAGGGCCAGTGGGGAGGGAAGATCTTCCAGGCATTTCCAGGAAACTCTCA

GGAGAAGGGTGTAGCTGTGAGCAGTCTGCAGCTGCTGCTCACTGCGGCTAAAGGCTGGGTGT

GCAGGCCAGTCAGCCAGTGAGGTGCCAACAGCAGGCACTACAGTCCACCCCTTGACTGCTCA

25   GACCTACTGCTTTCCACTTTAAGCTCTCTCCATCCAGGCACAGCTTCAGGGAAAACTTACAATT

GGAGAAACAGAGGGATGAACTACAATGCCCACTTCTGCATGTGATTGTAAGACTGTCACTGAT

ACTCACCATCATGCCCCATCCCCACCATCCATTCTAGTGTCCCCTTCCCCTTGGCTAACACTGC

TGGTCTAGGTGACTTCCCTAGAGCAGGAGCCAAACCCTTATCCCTGAGGCATCTGAATCCTGG

ATTCCTTTATCAGGCTATTGTTGTTGTAAGTTGTCCATTCCCAATTACAACTGGACATGAGACT

30   ACCAAGAAACACCCTGGCAAATCATCTGAGTGCAAGCCATATTCTTCCTGCTCCATTATGTAG

CGGTAGTCCTACCTCCTAATGACAAGGGTAAATTGCCACATTTTGCTCCTTGTGCCAGGATGG

TAATACCTTTCTCTACCTGCTTGGCTACTGGCACAAGGAAGCACAGCATGACCAGGAGGCAAT

TGTAGCTGTACATTTAGTGAATGTGTTAATGTATCACCTGGTGGAAGGACCCCCTCTGAGAAC

CAGGACTTCTAGACCCACAAAACCTAAAGTTGTGAATGGCGGAAGCACAAATTTCCCAAGTG

35   GATCATGGAGAGTGATGAAGAGTTCTTGGTTCCCAAACCCACATATTTTACCTTTCAGGAACA

TGGCCTCATCCCATAGCCATTAGAGTGCATATTGCATTCTGGAGGAGACTGGGCCCTCCTCAT
GGGTGTCATCTTCAAGATGACAGCTCCACTGTGCCTCCAAGAGGATGCTCCACCACCCTATCT
GTGATTCCTTGGTTAGCAGGACAGGCTGCTGCACTGAGGGTAGGAAAGGCAAGTCCATTGAT
GGCTGGAATACATGTCAATCCAAGTCAAGAGAAAATGCCGCCCTTTCCAGGTTGGAAGGGGC
5   CCGATTTAGCCAACTTGTCACCCAGTAGTGGCTGGTTGGTCTCCTCCAGGAGCAGTGTTATAC
CAGGAATTCAGCACCAGTCGCTATTGCTGGCAGTTCTTACATTCAACAGCAGCAAAACTAGGT
CAGCCTTGATGAGAGGGAATGTATGCTTCTGGGCACAGGCATGGCTTCCTTCTCTGACTCCAT
GACTATCTATTTCTGAGTGCATGGTGGCCGACATTCAGCTGCCTGCCCATCCTATCCACTTGGT
TATTATTGCCTCTTCCACAAGAAGTGGTTTCTGGCTGTCATTAATGTCTCATACTTTGTGCCCA
10 CTCACACAGGTTTAGCTCTACAACTTTTCCCCATGCCACCACTTTTCCACAATCTTCTAATGTT
GCTCCTTCCAAGCTACTGAAGAACGAGCTAAGCTATTCACCAATGTCCATGAGTCTATATTTA
CCTTAGGCCACATCTCTCTCCACACAAAGTGAATAAGCAGGTGCACCCTCCAAAACTCTACTA
AGAGGATTTCTTCTCCCCAGTGTCTTTCAGGGCCACCTTGAGTGGGGCTGAAGTACAGCAGAA
GTCCATTTCCAGCTTGCATCAACATTCCAAACTAACCTATCCATGATCAATGCATAGATGGGTT
15 TTTCCCTCCTCCAGCAGCTAGACAAAAGACACCCCCCACCAGGAGGCCATATTTGCATGTGGG
TGAAAGAGAGGCACAGGGGCCAATATTCGTGCAACAGTGGTAGATGGCAGGTGGGTCTGGGC
CACCTGTCCCTGCAGCTTATCTGTGCCATCTGGACCTGCTCAAGCCTGATTCCAGATATACCAT
TTCCATCTTATGATGGATGGCTTATGACCTAGTGGGTCTGACAGCACCAAACTCATAATGGGC
AGTTATGGCCACATGGTCACTTAATGTCCTATGGTCAGACACTCTGCTGAGTGGCATGCCAGG
20 AAATGCTTTACAAGTGGTGTTTGGTTCTCTGCTGCAGATGGCATGACCTTGGTCCGGAGCCCT
AGGGGTTTGGACAGTGACTCCTGTTGGGGCCTAATCTCACATTCCATGCAGAGTATCATCAGA
TTTGCCAATCACATAGCCTAAGGGTCAGGACTGATCCAACCAGTTTTTGCAGAGATCAAACTG
GAGAATGAAAGGTTGATATGATGTGACCATCATATCACGTTTTTCTCTCTTGAAAAGTATGCA
GATGTCTGAAAGAGACAAGTGCCCCAGGAGAAAATGCATGCCTTCCTCAGGATCGGCCCCCA
25 CCTCCCCTCCTGGCCACAAGGAGGGTCAAATCTCAGCATGGCCCAACTTGGACCTGTCAAGGA
AGAAGAAAAAATTGTATGCCAAAGGAACTCAGTCTTTGGCTAACAAGTACTAGACATCCTTT
AAGTCTTTGAGAATGGTAATAATTTCTGCCATCCCTCCAGATTTGTGTTTTTCTGTTTTGGCTG
GGTGGGAATGCAGCATTTTCACTTTGCCTTTGTTATTACAAATGTTGCTTATTCTATAAATCAA
GGAACCATTGTAAGGGCTCTTCTGATGGTTAAGTATATCCATTCCAATGATTTATTCGGGATCC
30 AAGGAAATGATTTCTGGGTGAATACACAGAACTAGTGGATCCAATTTGAGACATACCTGGGC
CAGAACTATATTTGTCGTCTTACCCCAATAAGCCTGCACTCTACTAGGACAGCCATGACAGCA
CTTTGGGACCCTAGATATAAGTGTGAATTGCTGGCTGGGCATGGTGGCTCACGCCTGTAATCC
CAGCATTTTGGGAGGCTGAGGCAGGTAGATCACCTGAGGTCAGGAGTTGAAGACCAGCCTGG
CCAACACGGTGAAACCCCATCTCTACTAAAAAATACAAAAATTAGCTGGGCGTGGTGGTGGG
35 TGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGGAGAATTGCTTGAACCCAGGAGGCG

21

GAGGTTGCAGTGAGCCAAAATCACACCACTGCACTCCAGCCTGGGTGACAGAGCGAGATTCC

ATCTCAAAAAAGAAAAAAAAAGTGTGAATTGCTATGAAATCACTATCAAAAGATCTGAGT

GTTACCCTTACTCAGTGTGGTCGAATATAAATAGCCATAGGTTCCTGTTATACACACTTGCTGT

GGTGCTACAGAGTCTTTCCTCATGGGAACCCAGTCCCTCTTTCAGTCAATGGGTTCTGGTTCGA

5   GAACTGGCTGAGGTTTGGAAACTGTGCCTTTCCATCATAACTTTCCACTGGGGTGACTGACCTT

GGCCTTCTGTTCATCCTTTCTAGCCCCTAAGAATCCAACACTCTATTAGCCTTCTCCTTAGACC

CCTATAAGCTAATCCCTTCTAGTTGTTAGTCTGACCTTGGTGCCCAATATGATAATTATTCCCA

CTTTGCTTCTGATATGCTTCTAAGTGCTGCCCCTGGTCTCTGCCCTTAAGTGATCTATCATCCCC

ACTGCCATTAGGGGGAGAAGCTCTGAAAAGAGTTGTCTCCCATCAACTCTGGTCTACAAAGG

10 ACAGCCCTACTGAGCCTCAGCCATGTGCCCGACACCAGCAGATTCTTTACAGCCTGGGAAGCA

GAGTGTCTTCCCTGCCTTTCCAGGGAACATAGCCAGCTTACAGGCTTTTTGATCTTATAGAGTA

GGTCAGTTATATTTTGCCCCATTTCTTTTATCCTTTTGATCACTTCCTCTTGGCCCACCATGTAA

ACTCAAGCATCCCTGCTTCATTTAATCGAGCTGTTGCTTTTTCTAAGCTACCAAGAGCAACCCC

AGCAATATATCAGAGCCCTCTCTTGGGACCCTTGCTAGGGTGTTAAATCCTGCATCATAGGAG

15 AATGCCCCCACATCAGCAAAGTCCCCTTATCCTCTTGATATCCCACCTGCCCCAGTCCAGCACC

TTCAGGATCTGGTCTCAATCACAGGATCCAGCACCTTTGGGACTGTTGCAAGCATAAGATCCA

GCACTTTTGGGATCTAGTCTCCCACTTCCTGCTAGTACTTGTTAGCCAAAGACTGAGTTCCTTT

GGCATACAATTTTTTTCTTCTTCCTTGACAGGTCCAAGTTGGGCCATGCTGAGATTTGACCCTC

CTTGTGGCCAGGAGGGGAGGTAGGGGCCGATCCTGAGGAAGGCACTCATTTCTCCTGGGGC

20 ACTTGTCTCTTTCAGACATCTGCATACTTTTCAAGAGAGAAAAGGCCTCCTTCTCACAGCAAG

ACTACTTCTGTAGATGCAGGTGGCTCGTGGGAATCTGGCAATTCAAAATTCTCAAGTGTACTC

ACTAGCACATTAGAAAACCAGTAGTACACATCTCTTTCCAAATCTTCATTCAGTGACACTATG

TCAGTAGCTGGAAATGGGCCATGGTGGGTGTATTTAAACCATGAAAATCAGAAAATGCTACA

AACCAGGGCATCCCGCATCTCTAGACAGCAGATTGTTGGCCATTTCCCAGCATACCATTGTGT

25 ATACTCCTTCCCATCAGGGCCGTGGCTTGCCTTGGTGGAGGACTCAGCCCTTGCTGAAGTTCTG

CTACTGCTCTTACAATTGAGTCCTATGCCTGGTCTCCAGCTCTGCCTGCCTCACTACAGGAGAC

AAGCATCTCTTTGAACACTGCCGAGAAGACCCTCTGGCTCTCAGGCTTGGCTTTAAATCGATA

GACCTGAGCCTGCCATTTTCTCTTTTCCATGCATCACTCCACTGATCCACAGGTCTCAGTGGCA

TAGTCCTTCGGGTTAGCATCTCCCCCACACCCTCGGTGCCAGAGACACTGAGTAAGAAAGTAC

30 CTCCCTGTCTACCCCCATCCCCGCTCCCCACAGGCAGGGCCTTGGCGATCCACTGCTGCAATGT

GCCAGAGACTGTCAGTACTCCTACCACCAGTGAGGTGGCAACCAGCTGGGAAGTGATCCAAC

TCCAGAGTCCCGCCCTCATAGGCTGATTTCTAGGACCACCCCTGGTATACTGTGTTAGGTTCTT

GAAGCAGAGCCTGAGATAAGGATTCTGGCACCTGTGATTGAGTGGGAGGGTGCTCTCAGGAT

GAGATGGGGTAGAAATAGGCAAAGGTACAGATTCAGCAGCAGTTGAGCCTCAGTCTGACCCA

35 GCAGGGAGCTCTCAAATGTGAATGACATCACAGAGTTGTCCCTCTGAGGCAGGGGCCAGCCTT

22

TGTGCTCCTACATGAGTCAGTCACTGGCTGGAGGCCCCTGGGGAAAGGCTAGGGCTGCCAGCT
TTAGCAAATAAAAAATTAGGGCACTCAGTTAAATTGAATTTCAGATAAACAACA

The genomic DNA or YKT6 SNARE gene is 39,000 base pairs in length and contains seven exons
5   (see Table 1 below for location of exons). As will be discussed in further detail below, the YKT6 SNARE
gene is situated in genomic clone AC006454 at nucleotides 36,001-75,000.

The human liver glucokinase is depicted in SEQ ID NO:2:

MPRPRSQLPQPNSQVEQILAEFQLQEEDLKKVMRRMQKEMDRGLRLETHEEASVKMLPTYVRSTP
EGSEVGDFLSLDLGGTNFRVMLVKVGEGEEGQWSVKTKHQTYSIPEDAMTGTAEMLFDYISECIS
10   DFLDKHQMKHKKLPLGFTFSFPVRHEDIDKGILLNWTKGFKASGAEGNNVVGLLRDAIKRRGDFE
MDVVAMVNDTVATMISCYYEDHQCEVGMIVGTGCNACYMEEMQNVELVEGDEGRMCVNTEW
GAFGDSGELDEFLLEYDRLVDESSANPGQQLYEKLIGGKYMGELVRLVLLRLVDENLLFHGEASE
QLRTRGAFETRFVSQVESDTGDRKQIYNILSTLGLRPSTTDCDIVRRACESVSTRAAHMCSAGLAG
VINRMRESRSEDVMRITVGVDGSVYKLHPSFKERFHASVRRLTPSCEITFIESEEGSRGAALVSAV
15   ACKKACMLGQ

and is encoded by the genomic DNA sequence shown in SEQ ID NO:6:

ACTAGCACATTAGAAAACCAGTAGTACACATCTCTTTCCAAATCTTCATTCAGTGACACTATG
TCAGTAGCTGGAAATGGGCCATGGTGGGTGTATTTAAACCATGAAAATCAGAAAATGCTACA
20   AACCAGGGCATCCCGCATCTCTAGA
AGCAGATTGTTGGCCATTTCCCAGCATACCATTGTGTATACTCCTTCCCATCAGGGCCGTGGCT
TGCCTTGGTGGAGGACTCAGCCCTTGCTGAAGTTCTGCTACTGCTCTTACAATTGAGTCCTATG
CCTGGTCTCCAGCTCTGCCTGCCTCACTACAGGAGACAAGCATCTCTTTGAACACTGCCGAGA
AGACCCTC TGGCTCTCAGGCTTGGCTTTAAATCGATAGACCTGAGCCTGCCATTTTCT
25   CTTTTCCATGCATCACTCCACTGATCCACAGGTCTCAGTGGCATAGTCCTTCGGGTTAGCATCT
CCCCCACACCCTCGGTGCCAGAGACACTGAGTAAGAAAGTACCTCCCTGTCTACCCCCATCCC
CGCTCCCCACAGGCAGGGCCTTGGCGATCCACTGCTGCAATGTGCCAGAGACTGTCAGTACTC
CTACCACCAGTGAGGTGGCAACCAGCTGGGAAGTGATCCAACTCCAGAGTCCCGCCCTCATA
GGCTGATTTCTAGGACCACCCCTGGTATACTGTGTTAGGTTCTTGAAGCAGAGCCTGAGATAA
30   GGATTCTGGCACCTGTGATTGAGTGGGAGGGTGCTCTCAGGATGAGATGGGGTAGAAATAGG
CAAAGGTACAGATTCAGCAGCAGTTGAGCCTCAGTCTGACCCAGCAGGGAGCTCTCAAATGT
GAATGACATCACAGAGTTGTCCCTCTGAGGCAGGGGCCAGCCTTTGTGCTCCTACATGAGTCA
GTCACTGGCTGGAGGCCCCTGGGGAAAGGCTAGGGCTGCCAGCTTTAGCAAATAAAAAATTA
GGGCACTCAGTTAAATTGAATTTCAGATAAACAACAAATTATTTTTTAGTATATGTCCCAAATT
35   GTGCATAACATAATGTGTTTTCTCCGCCAGCCCTGGGAAGGGCGTAACTTCCCAGGTATTTCT

23

AGGTGAAGTAACTTTGTAGATCAGGAGTAAGTCCCAGGAAAGAAGTCCAGCTCTTCTCT
TCAGCCCTGGGCAGCTGGGGGTAGGCACAGGGGCCCAGCAGGCACCCATA
GCATCTCCTACAGCATCTGAAATGAACAGGGTCATCACGTACTACATACAAATGTACCCACTG
CTGAGTTCTTCAGGGATTATATCATTAGGTACTTGGTATTTTAAATACATTACATTATGCAGAA

5   GTCCTTTGTGGATTGCTATATTTGGAGAGTTTTGTGATATTGGGGGGATTAGATGGAGTTTTCA
GATGGGCAT CATACGGTTTTTCATTTAAAACCCTAGAGTATTGTAATCCTAGGGAGTGA
TCCTGCGATTAGTAAATTAGCTCTCCAATAGATTTTCAATGTGGTTGCAAAGGACATGCATGT
GGTTCACCCTCCCAGGAAATCCAGAAGGGCAGCATTGGCCTGAGTGGCCTGAGTTTGGCTGGT
TGGGCTGGTAATGCTGGACAAAGA

10   CAATGGGTGGAATGGTTTGCTTCCCTCAGTCCTTTCAGACACAGCCCAGC
CCACCACGTCAAGCCAGTGGGTGCATCTGCAACCAATCCCCATGAGAACT
GCAGCCTCTCAGAGGTGGGCAAGTTGGCCCGGGTGGGTCAGGAGGATCAG
ATGTTGAGGAAATCTTTGGATTGGAGGCAGGCAGAGCAGGGAAGCATCGG
GTGATTCTATGACAGACCCAGGGCTCCAAGCTGCAGTTCAGGAGGGGCAC

15   TGGCACGGCCTCTGCTCAACTCCCCCTTGAGTGACATCAGGTGAAGTGCC
GACAACACAGAAGGCAGCAAATGCTGCCAGTCAGGTCTGCTTCCCAGGAC
AGCCAGTTGCTAACCCTTCTCCAGCACAGCACTGGATTTTGGTCACCTGG
CTGGGAGCTCCACCTCCCCAGCTGCTGCCTCACCTGCTTTTCCAAACCCC
ACCCTGTAAACGGTAACTACATTTTGTGCCCACTACGCCTCGTTTCCATC

20   TCTTTGGAGCACCTCTCACGTGGAGCTGAACAGAACGACCTGTTAAGCCC
ACCGTGTCTGTTAGGGTTGTCTAGGCTGTATCAGATACCCAACTAAAACT
GGATTCACCAACAGGTATTGTCAAAGCACATAAGAAAGAGTCCAGAGGCA
GGCAGCTCTCAGCCTGGTGTCAGGCTCTGGGTCAGCTTTCCAGATTCTCT
TAACCTTCCCCACATCTGCCAGATGCCGCCACAGGCACAGGAGGTACAAA

25   CAAACCCAAAAATGTTCTGGAAACAAGAAGGGAAGGGGATCCCCACCATA
TCTCCCCAGAGGCCTTCCTTCTCACATCTCACTGTACTGAAGCCAGCTCT
AGCAGAAGACAGCAGGGTGAATTTGTCCAGGGTATTCAGCCCCCAGTGCT
GGGTCCATTACTACTTGACCCCTGAATAAAACAGAGGTTCCATGAGCAAG
AAGGAAGGGGAACTGGATGTTAGAGGGCAAGAATGTATCCATCCCACCCC

30   TAGGAGCACGCATGGACAACTGCCCCATTTTTGCTCCTATTGCAGCCCAG
GGGCTAGCCCAGAGACCTTGCCAGTGCTGAGTCACAAGATGCTGGGAAAG
TGAGACCAGAGCCTGGTCTTGGGGAACAGCTCAAGGCCGCATTGGTCTGC
AGGTCATAGAGCAGCTGCTGAGCAGTGAGAGCCCACGATGGGCCAGGCCC
TGGGTCTTGGAGACCTGAATGAGATAGACTGGGTTCCTGTTCTCCTGGGC

35   ATTGCCTCTTAGAGGGCAAAGACAATTAACAATAAACAAATAGAACATGA

AGTGTTTTCCGATAGTGACTGATATACTTTGGATATTTGTCCTCTCCAAA

TCTCATGTTGAAATGTAATTCCTTATGTTGGAGGTGGGGCCTGGAAGGAG

GTGTCTGGGTCATGGGGGCAGATCCCTCATGAATGGTTTAGTGCCATCCC

CTTGGTGATGAGTGAGTTCACGTGAGAGCTGGTTGTTTGAAAGAGCCTGG

5   CCCCCTCTCATTCTCCTGCTCCCACTCTTGCATGAGACACCTGCTCCCCC

TTCTCCTTCTGCCATGATTTTAAGATTCCAGGGACTTCACAAGAAGCAAA

TGCTAACGCCATGCTTCTTGTTCTGTCTGCAAAACTGTAAGCCAATTAAA

CCTCTTTTCTTTGTAATTTATCCAGTCTTGGGTATTTCTTTATAACAGCA

CAAGAACAGCCTAATACAGTGATGCTCTCCAAGTGACCTTTGGGCTGAGA

10   CCTGAAGAAGAAGGGGAAGCAGTTAGGTCTGATAGCTCATGCCTGTAATC

CCAGCTCTTTAGGAGGCTGAAGTGGGAGGACTGCTTGAGCCTAGGAGTTG

AAGACCAGCTTGGAAAACATAGCAAGACCCTGGCTCTACAAAAATATTTT

TTAATTGGCCAGGTGTGGTGGTGCACACCTGTAGTCCCACCTACTTGGAA

GGCTGAGGCAGGAGCATCTCTTGAGCCCAGGAGGTTGAGACTGCAGTGAG

15   TCATGTTCACACCACTGCACTCCAGCTTGGGTGACAGAGCAAGACCTGTC

TCGAAAAAGAAGAAAGAAGAAAGTAGGAAGAAGAAGAAGAAGAAGAAGAA

GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAGGAAGA

GGAACAAGAACAAGAAGAAGAACAAGAAGAACAAGAAGAAGAACAAGGAG

AACAAGAAGAAGAATAAGAAGAAGAAGGAGAAGAAGAAGAAGGAGAGGAA

20   GAAGAAGAGGAAGAGGAGGAAGAGGAGGAGGAGGAAGATGAGGAGGA

GGAAGCAGAAGCAGAAGAAAAAGAAAGAAAAGAAAGAAAGAGAAAGAAAG

AAAAGGGAAGGAGGGAAGGAAGGAAGGAAGGAAAAAGGGAAGGAAAGGGA

AGGAGAGGGAGAGGGAGAAGGAAGAACAAAGAAGAAAGAAGGAGAAGCAG

AGGCTTGTGCTGGATAGCCTTGCTTTTGCCAATGACCTTGCTGATTTTCA

25   GGGGGTCCTGGTGTCTTAGTCCATTTGTGTTGCTGTAAAGGCATACCTGA

GGCTGGATAATTTACAGAGAAAGAGGTTTATTTGGCTGAGAGTTCTGCA

GGCTCTACAAGAAGCATGGCACCAATGCCTACTTCTGATGAGGGCCTCAG

TCTGCTTCCACTCATGGCAGAAGGTGAAGCAGAGCCTGCATGTGCAGATA

TCACATGGTGAGAGAGGAAGCACGAGGGGGCAGGGAGGTGCCAGCCTCTT

30   CCTAATAGTAAGCTGTCTTGAGAACTAATAGAGTAAGAAATAACTCACAC

CCTGCCCCCAAGGAAGGGCATTAATCTATTCATGAAGTATCTGCCCCCAT

GACCCAAACATCTCCCATTAGGCCCCCACCTCCAACATTGAGGATCAAA

TTTCAACATGAGGTTCCGGTGGGCAAACATCCAGCTATAATACTGGGCAA

TGCTGACCAGACTCTTCCCCTCTCAGGCCCAGAGCTCCTTGGCCCTGTAA

35   CAACAGAAAATTGCGTTTGAGTGTCAAGATTTTTCCTTTAGTCCCCATGC

AGCTCCTTAGAATGAGGTGGCATCTTCTCCCTTTTCATAGGTGAAGAAAC

AGAAGCTCTGGAGGAACGAATCATTCATCCAAGGTCAGGTAGCTAGTAAG

CGTCCCACCAGCTCCCCAGATCTCCTGTTTCCTGTCCCAAGTCCCACTGA

GTGAGCTGGAACAATGGCTTCACTGGCACCTGCCGGGAATGGTGGCAGGT

5   GCCTATAATCCCAGCTACTCGGGAGGCTGAGGCATGAGAATCACTTGAAC

CCGGGAGGCAGAGGTTGCAGCGAGCCAAGATCACACCACTGCACTCCAGC

CTGGATAACAAACGGAGATTCCATTTAAAAAAATTAACATATAATATACA

TACAGTAACATTCACTTTTTAAGTGTACAGTTTGATGAGTTTTATCAAAT

GTATATGGTTATATAACCACCATCACCATTAAGGCAGAATCTTCCCATCA

10   CTCAAATAATTCCCTCAGCCCCACCTCTTGCTGTCAATCACTTCTCCCAC

CCTAGCCACTGGAAATCATTCATCTGTTTTCTGTCCCCTTGGTTTTGCCT

TTTCTAGAATGTTCTATACATGAGACCACTGAGAATATAGTCTTCTGTGT

CTGGCTTCTTTCACTTAACATAATGCCTAGCTCAGCAGTGTGTCAATCCT

CCCTCCCTTGCCATTGCTGAGCAGTGAGTATTCCACTGTATGGCTGTGCT

15   ACGGTGTGTTCATCCATTTATTCATTCACCAGCTAATGGGCATTTGGATT

GTTTCCAGGCTTTGGCTATGATGAGTGAAGCTGCTGTGAATGTTCAAGTA

CAAGTCTTTGTGTAGACAGGGGTTTTCAATTGGCGGGATAAATACCTAGG

AGTAGTATCGTGTGGTTAAGCGTACGTTTAAACTTAGAAAAACTGTCAAA

CTGTTTTCCAATGTGGCCTGTACCATGTTGCATTTCCATCAGCAGTGTTT

20   GAGAATTCCAATTGCTCCACATCCTCCTCCCGACACTTGGTTTCACCCAT

CTTTTAAATATTAGCCACTCTGGTGACTGTGTAGTGATATGTCAGTGTGG

TTGTAATTTGCATTTCTATGATTGACTAATAATAATGTTGCAGATATTTC

TGTATGCTTAGTGGGCATTTTTGGTGAGTTTTTAAAAATTGGGTTGTTGT

CACCGTCTTATTGAGTTGGAAGAATTCTTTATATGTTCTGGATGTTTATT

25   CATGTGTGTGTCTGCTAAGAGGTGAGACTGGTTCTACCCTGGTCCTAACA

AGCACCCTGGGCCTGCATCCCTTTTTGTGTCTGTGAGCTGGGTCTGCAGC

CCTCTCCTCCCACTACCTACTGCCCAGCAGTACCCCTCACCCATCACTGT

GGCTCCTGCAATGACATCTCAGCCTGTCTCTCCCTCCCTCCAGCTAGCCA

GAGGCAGGATGGCTCAGTGACACAGGGTGGGCCCTGAAGACAGAGTGCCA

30   GGGTTTGGACCTTGTATTAGCAAGAGTCACAAGGGAAACTTACTTTATCT

CTCCATAGCTCTGTTGTGAGGATCCAATAAATTAATCCATAGAAGAGCTT

AGGACAGCACCTGGCACAAAGTATACATGAGCTATTATGATGTTATTCTT

CCAACCCATTGTTTCTGTGTTGTCATAAACATGAATGCAGGACTCAGTGT

CCCAGCTCTGTGTCCCTCGCATACATTCCCTAACAGCCCACAGGTCTTGC

35   CTGTCACCGCCTCATTCAATAAGTGATGACTCTGCCTCTTCCTTGGCTGG

GGCCTTGCATTGGACATTTCTGTATCCATATTTGTTTTTTAAAAACTAGC

TGTTGGCCGGGCGCGGTGGCTCACATCTCTAATCCCAGCACTTGGGAGGC

AGAGACAGGTGGATCATGAGGTCAGGAGTTCAAGGCCAGCCTGGCCAACA

TGGTGAAACCCCATCTGTACAAAAAATACGAAAATTAGCTGGGCGTGGTG

5   GCATGCACCTGTAATCCCAGCTACTTGGGAGGCTGAAGCAGGAGAATCGC

TTGAACCTGGGAGGCAGAGGTTGTAGTGAGCCAATATAGCGCCACTGCAC

TCCAGCCTGGGCAACACAGCAAAACTCCATCTCAAAAAAAAAAAAAACAA

AAAACAACCTAGCTGGACTTGACACTCTTGTTAGAGGAAGATTTTTCCAC

ATCTGTTAACTTTTCTTCTATTGTTATCCATCTGTGCAGGTTTTTCTGTC

10   CTCCTGAGTCATTTTGATAATTTATATTATATTTTGAAAATCATCCATTT

CCTATAGTTGTTTATTAGTGTCTTCTCTGTTATATTTGATCAGATTACCA

AATCTTGCTCATTGATTGCCCATTTATTTTATTGTGTTTATTTTTTTGAG

ACAGGGTCTCACTCGACAGCCCAGGCTGAAGTGCAGTGGTGCAATCATGG

CTCACTGCAGCCTTGACCTCCTGGGCTCAAGCAATTCTCCCACCTCAGCC

15   TCCTGAGTAGCTGGGACCTCAGGCACACGCCACCACAGCTGGCTAATATT

TTATTTATTTATTTATTTATTTATTTTTGTAGAGATGGGGTCTCACTATG

TTGCCCAGGCTGGTTTCAAACTCCTTGGTTCAAGTGATCCTCCTGCCTCA

GCTTCCCAAAGTACTGGGATTACAGGAGTGAGCCACCATGCCCAGCCCCT

ATTTACTTTATAGTAAGTGCCTTCATGGGCATAAATGTTCCTCTGAGACA

20   GCTTTGGCTATTAGCCATACTTTTAATATTTTGTACATTCATGGTTATTC

ATTTATAAATGGTCTGTAATGCAATGCAGATTTCCCCTTTGGCCCAAATG

CCATTTACAGCAGCACTTTTCTCTTTCTGAGCAGACAGAATATTTTGGTT

TCCCCTCTGTTGTTTATTTCTCGTCTGCCTCGCCTCATTTGCTAGGTGTT

CCCTTGGTGTGCCTTAAGTATGAGCCACTCAAATATTTGTGTTTCTCTAA

25   ACACCCCTGACACTGTCCTGCTGGTTTCTCTATCTGGAATATCCTTCCCT

TCTTGGCCAGTTCCCCCTAGTGCATCAAAGAAATCCTGCTCTTTTGCCTT

CAGAAAACAAAACAAAACGAAACCTATCAGTCTCCTTATGTCCCCAAAGA

CATAGCTTTGCTGGTATCTGGTTGTATTGAGCTGTTCATTTGTCTCTTCT

GCTAGATGGTAAGCTCCTTGGAAACTAAAAACTAATCACTTTTCTAACTT

30   CAGACTGAGCACAAATTAGGTTCTCAAGAAACATTGAATAATGAGTGATC

CGGTATCCCCTTCCAACATATTTTTGGTCATTGATACCATCATTCTGAGT

AGTTACTAGGGAACACTTCACTGCAGTAACCAATACAGCAAAACGTGAAA

TACAGTTACATAGTAGAATTGTATTTCTTGCCCATATAATAGTCAAGTGC

AGTTCTTCATCAGCTGGGAGGTTCTCCTCCACACAGTCATTTAGGAATCC

35   AGGGAACATAGCAGAGGTTGCTAGCTCTAGACCCAAACCCATGTCCTCTT

27

TGTCCACAGTGAGGACAATGCCAGCAACAGCTGGCCAGCTGTTCTGTAGT

TCTCAGCCTCCCTCGCAGTGAGATGTCTCCATGCAATTTCAGTGGAGCAA

CATATACCATTTCCATTTCCAGGTGTAGGCTCCTAAGAAGAGGGTGGCTT

CTTCATGTTCTTTCTCACCTTTCCGTAGGCTAGCTGCAGATAATGATGAG

5   GCTTTAGGGAGTGGGTGGAGCCATAAAGTAGAAGCCTGGATTCCTAAATG

ACGGTGTGAAGTGTTCCCTAATTTCACGTAATTGTTTCTTAATTTCCTGT

TTGGGTTATTTGTTGCTAAGGTATAAAAAAACCCTGATTTTTGTGTGTTG

ATATTTGTGTGCTGCAACTTTGCTGAATTAGCTTATTAGCTCAATTTGAT

CTCAGATATTAGCTCAAATATTTTGGGAGATTATTTATGGTTATCTACAT

10   AAGATCATGTCATCTGAAATAAAGATAGTTCTATTTCCTTCTTTCTATCT

TAGTCCATTTGGGCTGCTGTAACAAAATGCCATAAATTGGAGGCTGAGAA

GTCCAAGATCAAGGCCCAAGCTAATTCACTGTCTGATGAAGGCCTGCTTT

CTGGTTCATACATGGCACCTTCTAGCTGTGTCCTCACATGGTGGAAAAGG

CAAGGTAGCTCTCTGGGATTCCTTTTTGTTTGTTTGTTTGTTTTGTTGTT

15   TTTGTTTGATTTTTTGAGACAGAGTCTCACTCTGTCACCAGGCTGGAGTG

CAGTGGCACAATCTCGGCTCATTGCAACCTCTGACTCCCTGGTTCAAACG

ATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGTACCCATCAC

CATGTCCAGCTACTTTTTGTATTTTTAGTAGAGACAGGGTTTCACCATGT

TGGCCAGGATGGTCTCGATCTCTTGACCTCGTGATCTGCCCACCTTGGCC

20   TCCCAAAGTGCTGGGATTACAGGCATGAGCCACCGTGCCTGTCCTCCGGT

ATTCTTTTTATAAGGGCTCTTTTTCTTTTTATGTGGGCTCTACCCTCATG

ACCTAGCACCTTCTAAGGCCCCACCTCTTAATATCATCACACAGCAGATT

TAATATATGAATTTTGAGGGGACACATTCTTTCCATAGCACTTTCCAGTA

TGGATACCTTTTATTTATTTTTCTTCCCTAATTGCTTTGGTTAGAAATGT

25   CTTCCCTAATTGCTCCACTACTATGTTGAAAAGAAGTGGCAAAAGTGGGT

ATTCTTGTCTTGCTCCTCTCTTAGGAAGAAAGTTTAAGTCTTTTGCCATT

AAATATGACGTTAGCTATGGGGTTTTCATATATGACATTTATCATGTTGA

GGAAATTTTCTTCTTGTTTCAATGATGACAGGGTGTTGAGTTTTGTCAGA

TGCTTTTTCTGCATCAATCAATATGACCATGTAGTTTCTTTGTTTTATTC

30   CATTATTGTAGTACATTACATTAATTTTTGCATGTTGAACTATTCTTGTG

TTCCTGGGATAAATTTCACTTGGTTATGGTGTATAATCCATAACCATAAC

CTGAAGATATGCTGAAGAGGCTAAGTGCCATGGCTCATGCCTGTAATTCC

AACACTTTGGGAGGCTGGTGTGGGAGGATCACCTGAAATCAGGAGTTTTA

GAAGAGCCTGGGCAAGTAAACAAGATCCCATCTCTACAAAAAATTGAAAA

35   TTACCGCTGGGCATGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGTGG

CCGAGGCAGGCAGATCACCTGAGGTCGGGAGTTCTAGACCAGCCTGACCA

ACATAGAAAAACCCCGTCTCTACTGAAAATACAGAATTAGCCAGGCGTGG

TGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAAAATC

ACTTGAACCTGGGAGACGGAGGTTGCAGCGAGCCAAGATCATGCCATTGC

5   ACTCCAGCCTGGGCAACAAGAGCAAATCTCCGTCTCAAAAAAAAAAAAAA

GAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAATTAGCTTGATGTGG

TGGTTGTGCACCTTTAGTCCTAGCTACTCAGGAGGCTGAGGCAGGAGGAT

TGTTTGAGCCCAGGAGGTTGAGGCTGCAGTGAGCCATGATTGCACCACTG

CACTCCAGCCTGAGCAACAAAGTAAGACCTCATCACTAAAAACAAATTTT

10   TTAATACTGAAGAATTTTATTTGCTGGTATTTTGTTGAGGATTTTGCATC

TATATTCACAAGAAATATTACTCTGTAGTTTTTCTTCTTGTAGTATCTTT

GTCTGGTTTCAGTATCAAGGCAATGCTGGCCTCATGAGATCAATCAGGAA

GTGTTACTTCCTCTTTTATTTTTTGGAAGAATTTGAGAGAATTGGTGTTA

ATTCTTCTTTAAATGGTTGGTAGAATTACCAGTGTAGACATCTGGTCCTG

15   GGATTTTCTTTGTTGGGAGGTTTTTTAGTACTAATTCCATTTCCTTACTT

GTTATTAGTCTAATGAGATTTTCTGTTTCTTCTTGAGCTAGTTGTAGTAG

CTCATGTGTGGAATTTTTCTATTTCATCTAAGTTATCCAAGTTTACCTAA

GTTAAAGTTCCATTTTATCTAACTTGGGTAAGCCAACAAACAATACTAAA

TTGTTCATAGTATTCTCTCATAGTCCTTTTTTTCTCTAAAGTCAGTAATA

20   ACGTTCACTCTTTCATTTTTTCATTCCTGATTTTAATAATCTGAGTTCTT

TCTCTCCCCCTCCCTGCAATTGAGAGTCATTTAAAAGTGTCTTGATTAAA

TTTTATATATCTGTGAGTTTTCCAGTTTTCCCTCTGTTATTCTCTTCTAG

TTTTATTTCATGTGATCCAAAAAGATACTTTATATGATTTCAATTTTTTT

ACATTTACTAAGACTTGTTTTGTGACTAAAATATCCTTGAGAATTTCCAT

25   GCACATTTGAGAAAAATGCACATTCTGCTGTTGTTGGACAGAGTGTTCTG

TATATGTCTGTTAGGTCTAATTGGTTTAGAGTATTGTTCTAGTCCTCTCT

TTCCTTATTGATCTTCTGTCTAGTTGTTTAATCCATTATTCAAAGTAGTG

GCCGGGCACGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCGAG

GAGGGTGGATCACAATGTCAGGAGGTTGAGACCAGCCTGGCCAACATGGT

30   GAAACTCCGTCTCTACTGAAAATACAAAAAATTTGCTGGACATGGTGGCA

CACGCCTGTAATCCCAGCTACTCAGGAGGCCAAGGCAGGAGAATCACTTG

AACCCAGGAGGCAGAAGTTGCAGTGAGCTGAGATCGCACCATTGCACTGC

AGCCTGGGCAACAGAGCAAGACTCTGTCTCGAGAAACAACAAAAACAAAA

ACAAAAAACAAAGTAGTGTACTAAAGTCTCCAACTACTATTGTAGAACTC

35   TATTTCTCCCTTCAATGTTGCAAAATTTTGTTTCATGTATTTTGGTGTTC

29

TGTTCTTTATAATTTTTATATCTTCTTAATGGATGAAAACTTTTATCAAC

ATATAATGTTCTTTGTCTCTTGAGACTTTTTTTTTTAACTTAAAATCTAT

TTGGGCTGATAATACAGCCACCACAACTCTCATATTGGTTGTTATTTTCA

TAGAATATCTTCTTCCATCCTTCTACTTTAAAATTCTTCTATCTTTATAT

5   CTAAAGTGAGCCTCTTGTAGATAGCATATAGGTGGATAATGTTCTCTTTA

TTCACTCTGCCAATATCTGCCTTTTAACTGGAGTTTAATCTATTTATATA

TAAAATAATTACTGATTAGGAAGGACTTACTTCTACCACTCAGCTATTTT

TTTTCTGTGTGTCTTATACATTTTTAAGTTTCTCAATTCCTCCATTACTG

GATTTTTTTTTTTACTTCTTGATTTTGTGTCTGTGTTGTTACATTTTGAT

10   TATTTTCTCCTTTTGATAGCGGCAGGAGGCAGCCAAATGCCTGGCAGATA

GAAGCTTGTCCCCCATGAAACCCCACCTTCAAGCCAAAAAATAGCCTGAA

GGCTGAAAGACCGGACTGCTGGTCCCAGATGAAACCCATGATCCAGAGTG

AGAACTTCCATTCCTGTTTGCCTGCCCTCTAAATAATCCCTTTTAACCAA

TCGAATGTTGCCTTTTCCAATACTACCTATGGCCTGCCCCTCCCCCATTC

15   TGAGCCCATAAAAGCCCTGGAATCAGCCACATTGGGGGCACTTTGCCAAC

TTCAGGTAGGGGGACCACCTCTGTATCCCTTCTCTGCTGAAAGCTGTTTT

CATCACTCAATGAAACTCTCACCTTGCTCCCTCTTTGATTGTCAGCGTAT

CCTCATTTTTCTTGGGTGTGGTACAAGAACTCGGGAACCAGTGCACAAGC

CAGACTTGGTCTGGGCAGCACGGGTTAGTGGGCCATCTCCCACAGCAGGT

20   AGCATGGCCAAGTGAGGCCTGGGCAGGGCATCACCAAGGTCCCTGGCTTG

CAAAGTGACCAAGGAAAAAATCCTGTGTCACTTTCCTTTTCTCATATTTT

TTAGTTATTTTCCTAATGATTGCCTTGAGGATGGCAATTAACATCTTACA

CTTATAAGAAGCTAGTTTGAATAATAGTTCCAATAGTACATGAACACTCT

ACTCCTATATATCTCCATCCTTCTTCCTTTATATTGTTATTCCCACAAAT

25   TATGTTTTTATACATTATATCCTCACTAACATAAACTTATTATTATTTTC

TGCATTTGCCTTTTAAATCATACAGGAAAACAAGAATCACAAAGAAAAAC

TACATTAATATTTGCTGTTATATTTACCTATATAGTGACATTTAACAGTG

TATTTTTATGTCTTCAGATGTCTTTGAATTACTACTTAGTGTCTTTTCAT

TTTAGCCTCAATGTTTCCCTTTAGCATTTCCTATAGGGCAGGCCTGCCGG

30   TAATTAATTCCCTTTGGTTTTCTTTATCTGAAATGTCTAATTTCTTTTTT

ATTCTTGAAGAATAGTTTTGCTGGCTATAAGATTCTTAGTTAATAGTTTT

TTTCCCAGCACTTCAATTATTATTAAAGTGTTATTATTATTATTATTATT

ATTTTGAGATGGAGTCTCCCTCTGTCACTCAGGCTGGAGTGCAGTGGCGC

AATCTCTGCTCACTGCAACCTCCGCCTCCCAGGTTCAAGCAATTCTCCTG

35   CCTCAGCCTCCCGAGTTAGCTGGGATTACAGGTGCCCGCCACCATGCCCA

GCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTGGTCAGG

CTGATCTTGAACTCCTGACCTCAAGTGATACACCCACCTTGGCCTCCCAA

AGTGCTGGGATTAGAGGCATGAGCCACCATGCCTGGTCTAAAGTGTAATT

ATTATTACAGCTGCCATTTGGCCTCCTTGGTTTCTAATGAGAAATCATCT

5 GTTAAACTTATTGCAAATCCTTGGTATGTATGCTATGTGTCATTTCTCTC

TTGCTGCTTCCAAGATTCTCTCTGTCTTTGTCTTTTGACAATTTGACT

ATAATGTGTTTCAGTGTGAATTTCTTAGAGTTTATCCCACTTGGATTTCA

TTGAGCTTCTTGGATGTGTACGTTTGTCTTTCACCAAATCTGGGAAATTA

TTTCACCATTTCTCAAATATCTTTTTCTTCCCCTTTCCATCTCTCTTCTT

10 CTGGAGCTCCCGTATACTTAGTTGGCATGACTGATGGTATCCTACTGGTC

CCTCAGGTTCTGTTCATTTTTCTTCTTTCTTTTTTCTGCTCTGCAGACT

GGATAACTTCAATCGCCTTTTCTTCAAGTTCAATGATTATTTCTTCTGCC

TGCTCAAATTGGCCATTTAACCCCTCCAGTGACTTTTTCATTTCAGTATT

GTACTTTTCAGATCCAGAATTTCTATTTGGTTCCTCTTTAATAAATTCTT

15 TTTATTGTCATTCCCCATCTGTTCATACATTGCTCTCCCAATTTCCTGTA

GTTCTTTGTCCATGGTTTTCTTTAGTTAATTAAGCATATTTAAGACAGTT

GACTTAATGTCTTTGACTAGTAATTCCAATGTCTAAAATTCCTTATGGAT

AGCTTCTTTTAAATTATTTTTGTCCTGTTAGAGAGTCATATCTTCCTCTT

TATTTGCTTTGTAATACTTTGTTGAAAACTTAACATTTTGAGTAGTAAAA

20 TGTGGTAATTCTGAAGCCAGATTCTCCCCCTCCTTTGAGATTGGTTTTGT

TGTTTGTTGAGGGCTGCAGTTGTCCATTTGTATAGTGACTTTTCCAAACG

ATTTTTGCAAAGTATGTATTCTCTCTTGTGTCTGGTCACTGACGTTTCTG

TTCTGGTGCCTCTGCAGTCAGCCTATGACCTGGAAGAGCATTCCTTAAAT

GCATAGATTTTTTTAAAACCCAAGAAACAAAAAACCTAGCATGTATGTAC

25 CTTTTTAAAAATCTTCTGATAGATGCCACCTGGAAGGCTGCTGCTGCCTG

AAGGGGCAGAAACAAAGGCAAGCTCTACTCTGAGCCCTCAGGGAACCACC

AGATAAACAAAAGAAATTTGATTCTCCAAATTTCTGGAAGACAAGGTCCT

TTCTGCCCACTCCTGCTCCAGCCAGCTGCTCTAGGAACACAATTACTGTC

CACATGGCCACAGGAATGTTGAAGAATGCAGGATGGTAGCTGGTTTGCCC

30 ACACCACTCACTTATGAGCCATCAGCATGCCTCTCCCTTCATCGAGCACT

CCCATGGTTGCTGTAAGTGTCCAATCAGGTTCCAGAATTCTGAAAGAGTT

GACTCTTACAGGATTTTTTTCTTTTCTAACTTGCTGGTTGTTTAGATAGA

GGAACCAATTCCTGAAGTTTCCTACGTTGCCAGCTTCATGAGGATCATTC

CCTAGTAACTCTTTTCAGACAAAAAGCTTCATTGATTTACTGTAGGACTA

35 GCATCAAAGAGTCTATGCCACCTAGTCTGTCTCCTTAAAACACAGAAATA

ATCAGTATGCATTGGGGTAGGAGTTTGGCATTAGATCTGCCGTAAATCAA

GAGCTGGGGACAGCCCATGTCTTAAACTCTGACCCAAGGGCTAAAATATC

CTTTGGTAGCAACAACAGCTACAAACTATTGAACAACTTGTATGTGCCAA

GAGCCTTACCTGCATTATCCCATTGAATCCTCTCAACAGCCCTGTGAGGT

5 AGTAGAATTGTTGCCTGCCCCTTACTGAGGCCTAGAAACATTAAGGAATT

TGCCCGAGGCCCTAGAGCCAGTGAGTGGCAAAGCCAGTCTCCAGACTCAG

GCTGGAGATCCTACAGTTCTGTGTTACCCCAGTGTTATCCTGCCTCTCAG

CACAGAGTCTTGGATGATTCTCCTAACCCCTCCCTAGGCAATGCACAGGG

CTGCTCCCTGCACCCTTACTCATGCTCTGCTCTTCAACCCCAACAGTGCT

10 GGCCTTAGGCTTTATCCCTGACACCCAGCCCCAGGCTCCATTCCATCTGT

TGACAGAGGCAAACACTGGGGCAAAACTGACCTCTGTGGATACCACTGTG

TCCACCTCCACCAGCTTCAGCTGAAGCCTCTGAACATCTCCAGCATGGAA

GAAGCCCCAAAGGATATTTCCTGTCCCCCAGCATATGCTTGACCCTGAAG

CCCTCCCCATCTAGTCAAGAAGACCAAACTGTTAACAATCCTGGAGTCAG

15 AGTGACCCATGGGTGAATCTTAGCCAAGTCACTCATAGCTGTTGCATCCT

AGTAAATCCCTTAACTCCCATAGGCTTCAGTTTCCCTGCATATAAAATGA

CAGCCTTCAGCTCATCGGCCAGTTTCAATCCATCTAAAGGGTCTAGCACA

TCCCCTGGCATGTGGAAGCCACAGGGCACACACTAGTTGTGGTCATTTGA

TCCTGGCATGCTCTGCTGTCTCTCGGCTCTCCCCTTGCCTCTTTCCCTGA

20 TGTCCTGGCCATCAGCCACTGCCTAACACCCTCCCACTCACCAGGCCCTT

AGCCTGCCCCTTAGCACAAGAGCACAGCCGGTCTCAAGTCTACCCTGCTG

TAAGCAAACACTTGCAACATCATGCTGACCTCCAGGCCCTGTTGCATCAG

CGTGCCCACACTTGGTGCCCAGCTGGTACTGAGGGTATCAGGGAACAGGC

CAGTGGTGGAAGGGCGGACACTTTGGGTTCCCTGGTTTCCTGGCTCCCAA

25 TATCTTTCCCAATGGCATATGGGGTCTAGCAGCTTGGCTCATTTAACTGT

GAACCTCTACCCTTTAGAATCTGGGCCTCCAGGCTTGCTTCTGTGCAAAA

TGGCAGATAAGGCTCAACCTTTCTTTTTTTAACTTCATTGTTAAATATTA

CTCCATTAATACCCATTTACTGCAGAAAAGGTAGGAAATACAGATAAGCA

AAAAGGAAAATAAATTAAAATCCTCATACCACCATCATCAAGATAATTAC

30 TGTCACCATTTTGGTATATTTCCTCCCAATACATATATTATCTATATCGT

ATATACGACAAAAATGGATCATACTATGTTTCCTGTTCTTCCCCTGTGTT

AGTCATCTATTGCTGTATAACAAACTGCCTCAAAACTTAGTGGCTTCACC

TTTCCGTGTATTATGATGACAAGAATGTGGTATGACACTGTCTTATATCT

GGATCATATGCTAAAAGATAGAAAATGGTTTCTAAACTTATTTGTTCTGT

35 AATAACAAAATTTTATTTCATAAAGTGTTTTTAAAAAAAACCATAGTAGC

32

TTGAAACAACAAACCTTTGTTATCTCACACAGTTTCTGTAGGTCAAGAATTCAGAAGCAGCTT
AGCTGGGTGGTCTGGCTTGGTGTCTCTCCTGAGGTCAGGGTTTTGGCTGGGGCTGCATCACCT
GAAGGCTTGACTGGGGCCAGAGGA
CTGCTTCCAAAGTGGTCCACTCACATGGCTGGCAAGTTGGAGTTGCGTATTGGCAAGAGACTT
CGCTTCTTCTCAATGGATCTTCCCAGAGTTCTTGTAGGCAACCTCATAGCATAGCAGTTGGCTT
CCCCCAGAGGGAACAGTCCAGGAGAGAACAAGGCAGAAACCACAGGGTCTTTTCTGGCTTAG
GCTCCAAAGT CATACTCCACCATTTCTGCATTATCATATTAGTTACACAGGCTAGACCTA
TTCTGCATGGAAGAGACTATACCATGGGGTGAATACCAGAAGCAGGGCTATTGAAGGCCAGC
TTCAAGGGCGGCTACACATTCCCTTTCAACAGTATGTCATGAACATCTTTCCATGCCAATAGA
GCAGATGAATCTTACCATTTTTAATGACTACATGTAAGTGTAGCATAATTTATTTAACCAACCT
CCTGTAGTTG GGTATGTGGGTTGTGTCTCGTTTTTTGATAGTAGAATTAATCATCTTGAA
TATCCATCACCAAACTTGTCATATTATTTTCTTTTGATGAATGAAAAGAAAATCAAGTCATGT
CTGTCAATCAGAACCCTGAGCAACTAAGAAATGGGGGTACCACTGGGACATAGAGCAAGGTC
CCTTCTGATTCTGCTCTTGTCTTTCTCTCCCCATGAAATGGGGAGTTCACTATCTACTGAGACA
TCCTAGCCCA CAGCTGCACAGTTCTGTCTTTTTAGAAAGCTCTAAGCAGAAACAATGTTC
ATCCATCCTCCTCGGGACAGCCCTTGAGCTACTGAAGACTCTAAGCATGTCCTGGTCATCCTCC
ATGAGCCATCATCTCTGAGGCCCTCCCCTTCTTGGCCCCTCTTCTCTGGACAGGTTCTGGACAG
TCTTGCCCTTCCAAAATTCCTGAAAGCAGGAACTGTTCCTGCTACAATGACTCTCAACTCCAGT
GCAGTAC AGACTGTTGGTGTCACCCCTTATCCTGAAGAAGAGGCACTGAGACAGGAC
AAGGGTGGGTGCCCAGGAGGGCTGGCATGAGTCATGAGAATCTGGTCCCGGAGAATTAGACG
GTGTGGGGAAGTAGGGGTGTTGGGCCGCTTTCTGGCCTCATGGATGCCAATGAATATCAGCAG
GTGGCTCCCAGAAAGGAACTCTAGGGGATGCCTGTTGCTCTAAATAGAGGCTAGAGAGGGCA
CTGGCAGTTCAGT CAACCAAGAAAGGGGGCCCACTTGCCTCAGCTTCAGGCTTTGTACACATC
CTCAGCCTTTCTTGAGAACTGAATTTAGATTCTCCTCCCCTGTGCTGTGTGCTTGGCCCAGAAG
AAGGGCAAGTCTCGCTGGGTGGCTGCTTCTTGGCCTGGCTGAACCAGAAGGCCCCAGTGCCAC
TCCAAACCTGGGTGTGAGCCCTGCCCCCATGAGCAAACAGTAGCTCAGAGCTGGGGGCTGTG
GGGGTCAGTGG CCTGTCACATGAGATCTGATGAGGCCATCTCTGCTCTATATTGGGAAAGG
GATCAATTGTATCAAGGGCTTTCTTGGGAGTGATCACTCTGGCCATTGGCGAGAGACCTGGCA
TTCTGACAAGGCACCCTCCATACCCTGACCCACTTGCCAGCTCCAGCTAATTTTAGCAGGCTTT
GGCAGGTGCCAGCAAGTACATAGCATGTGGATGTCACTCCCAGGTGAGCCCAAGGAGAGGCC
TGGGCCAGAGC CTGGAAGTCATGGTCTATGCCCATGGAGGCACCCAAAGCAAGCCTGAGGC
CTGGACTTTGCAGTCACAAAATTAAGAATGATACCCCTGTTTTTTGTTTG
TTTTGATCAGTTGGCCACCTTCCTCCACCACCCCTTCCCCAAGTTCCAT
CAGACCCCTGGATTGTATGAAATGCAAATCGAACCTCTCTGCAGATGAA
AATCCACTGGGGATCCCCTTGCCTCCAAGAGCAAGTCCAGACCTGCACCA

5

10

15

20

25

30

35

33

GCGCGGGCCAGGCCCCCTTAGGACCCCCTCCCTGTCCAAGGGCATTTCAG

TAAGTGTTCTGTGGCCAAGGCAGCCTGGTGACTTTCTGCCCGCACAAGGC

TGAGGAATGGAAGATGGGTAGGCTGGCTCTGCACACCCCTCCTGCTGGG

CAGCAATCCCTACCCCATGTTCACAGAGTGTGGCCGGCTGCCCCATGGCT

5　CTGTCCCCGTGGCCCTGTCAACTGTTACCCACATGGCCTACCCTCCCTTT

CTGCCCTGCCTCTGACCCCATGGCAGGGGGCAGAGTATTTGAGCAGCCGC

CAGGCTGAGCCCTTTCAGTGCAGAAGCCCTGGGCTGCCAGCCTCAGGCAG

CTCTCCATCCAAGCAGCCGTTGCTGCCACAGGCGGGCCTTACGCTCCAAG

GCTACAGCATGTGCTAGGCCTCAGCAGGCAGGAGCATCTCTGCCTCCCAA

10　AGCATCTACCTCTTAGCCCCTCGGAGAGATGGCGATGGATGTCACAAGGA

GCCAGGCCCAGACAGCCTTGACTCTGGTAAGGGTCACACCAAAGTTAGGG

ACTTTGCACTGGGAGAGCAGCACCCAGGGCAGGGCCCTTGGTTTTGCAGA

TTACCAAAACTAAGGCTGGGGGCAGGGAAGGCGAGCAGGCTTGGGGCACC

TTGGAAGGAGGCACATGGGCCTTGGGGGTCCTGGCTAGGGCAGCTGTGCCTGCCACTGGCCCT

15　CTGCCCACCACCCCTCCTCACTGTGGCTATCCAGTGTCCAGCCTCTCGAGGGGTTCTAGGGTAC

TTATTCCTGGAGCTAACGGTGACCCAGGACACCAGTGTCCGGGGCCTGGCCTGGGGCTTTTAT

GGGGGGAGCT GGCTGGCTGCCCAGGGCTGTCTGGCTCTCTGGGGGCTCTGCATGGCATTT

CCAGGGGTTGGTGGATCAGGGATTCTGTCCCTCAGGAGAATGTGGGCACTAGCCCAAGGCCA

CTCACTTCTGTGTACATAGCCACCTGAGGGCCCAGGAATGGAGGGGGCCAGGCTACAGCTGG

20　ACATCTGGCACTCGGATGGGCTCTGGAGCCCCCAGGCCTGCAGCATCTGCCCAGGGACTGCCC

TGGCCCTTGGCCA TTTCCTCAGGGACCCACAGCTCCACCAGCCGGCCCCTCCCAGTGCTGGAA

TAGACAGTTCCTCAGTCCACATCTGCCAAAGGCGGCACTAGAAGGCATCCTGCCTTTTTTACT

GCGTTCTGGAGGTGGGGTCACAAAGCACTGCTCACTGCATAAAAGGGACAGCATCCTGCCCCT

GGCAGCCCTGCCTGACCAGCTCCGCCTCTCCCACTGCTATCCAACCTGTACACCCTGGTGACC

25　ATGTCCAGGCC AGTGGCCTTAAGGACTGTCTCTGTACTGATGGCTCCACATCTACCTCTCC

AGCCAGACTCTCCTCTGAACTCGGGCCTCACATGGCCAACTGCTACTTGGAACAAATCGCCCC

TTGGCTGGCAGATGTGTTAACATGCCCAGACCAAGATCCCAACTCCCACAACCCAACTCCCAG

GTCAGATGGAACCTCTTCTTCCCAGGCCCTTCTGTTCCTCTCCTCAGCCCCTCCCACCTCCCTTC

AGAATAAGT CTAGACTCTTATCGCTTTCACCAAGCCTGCGCCCAGCATCCCTGCACAGG

30　GATTGTTAGGACAGCCTGACGCCCTGCTTCCACCCTGCCCCAAGATGCCCCTGCTCTGCAGCC

CGGCGCCTCCAGGCTTCTCACCTCCTGCTGCTCACAGCTCAGCCTCACTCCCTCCCTCCCCGCC

TCTGCTCCAGCCTCAGTGCAGGT

CCCTGCTCCCATCTTCTGGCAGCAGCTGCCCGACCTGGTCCCTCTTCATCTGTCCCCATTCCTTC

ACCCCCCAGCCTGTCCCCAACTTGACTGAGGTTCTTTCCTGCAGATCCCCGCCCTTGAGAGGG

35　GTTGGTCCCACTGTCAACTCTGCTTCTGTGCCCTGTGCCGCACCTGGCATTCAGTGAGCATCTG

CTGAAGA GATGAGGGTCAGATGCCCTGCAGGGAGTGTGGGGGCGTCCTCAGGCAAGA

AAAGTTGTACGTTTGGCTGTGGGCCCTGATTATGTGTCCTGTGACCTCTTGGGTGAGGTCAGC

AAGAGAAACCTCTGCAAGCTGGCTGGGGCTGCCTCCCAGAGGCTGCCAGGGGGAGGGACAGG

CTCTGTCTGTGCTCTTCTTCCGAGGCTACACCTGGGGCGCCAGGCTCTCAGGGCTCCCCAGGTA

5   CCACCACATTT CCTACACTGCTTGGGAAAGCCCTGTAAGTTTGCACAGACACCCAGCATGA

GGCTCGCCAGAGAGATACTTGTAGCTGGGGTCTGGGCACCAGGAACAGCTTGGTGCTGGGCC

TGAAGTCGGGCAGGATGCAGCCTGGCCAGGTGAGAGGAAAGCTTGGAGCCAGTGCCTGGGTT

CAAACTCCTCTGTGGCCTATGGTTCTGTGGGCTTGGGGAAGGGTTTGTACCTCTGTGTCCAGTT

TCCTCACTTATA AAAAAAGGAGATAATAAAAGTACCCATGTCCCAGGGTGGCTGTAGCAATA

10   ATAGGGAGGGGTGCCCAGAGCAGGTCTGGCACACAGGAAGTGTGCATCAG

CCTCAGTCCCTGCCATTGGGCTTGTCCTGGGAGTCTGTGAAGCCAACCTC

TGCTCCACAATGTGACCCCCAGGCTTGTGAGACCAAGCTGGGTCAGAGCT

TCCTCCTCTGGGGTTGCACCAGGAGGGGAACTTCTGCAGGCCCAGATGCA

CCCTGAGGAAAGGGCTTGTTCCCACCAAGAACAAGGCTCACCTTTGGAGG

15   ATGCTCCCCACATGAGAGGTGAACCCCCAGGTCTACTGGTGACTGCAGCC

TCGGAAGCTGACAGCATCTATCCTCCAACCCATGCCCACTGGGAAGTGTG

TGAGGGGTCCTCATAGGCCCTGCGGTGTGGACAATGCAGAGACCCTGTAG

CATCTGGCTAGGGCGGGGCCCAGATAAGAGCCCTGTGCCAGGAGAGCCTG

GCCGGTTCTGCCACTGTGGGGAGACAGGCTCCCCCACCCCATGTCCCCTG

20   CTTCCCTGCAGCCCACAGAGAATACAGACCTACTTTTACAGAAATCCAGA

TTTTTGTGTAAAAGTGTCTCTATTTTAAGTAGATTTTAAGTGGTGGCAGC

AAATTTAAGCTTTTGAGAATATTATACAGAACAAATCAGATTCACAGGCC

AGATGCAACTTTATTTACAGAAATGGGATCAGGTCCTACCTCAGGTCCCA

TCTCACGTTTTCACTTATGCCTATACGTCTCCTTCACGGGAAAGGCCACA

25   AGAGGCCCTGCGGTAAGTGTCCCGGTGTTGATTTAAAGTCCCCAACAGTG

AATATGAGGGTCCTCACTGTTGCAGCAAGAGGATACCCCCCTGTGTATCT

TGGAAATGCCTGCAGCCCTCTTGCTGCAGAACAGATTCTTAGGAGAGAAA

CTGTCAGATCAAAGTTAAACTTAGAGAAACTCCAAATTGCCCTCTGAACA

GACGGTATCAGTTTGACATCATCCAATACCGGGATTCCTCGGGGAGAACT

30   TTCTGGCCTAGAAGGCAGTAGAGCCAGGACTTCACCCAGTCAGTGGCAGG

GCCACACGTGGGCCTTGATACAGAGGGGGAAGACTTGAGCCTCCTCGACA

CCCTACAGGGCCCAGCCTCCCAACATGTGATAAGAGAAACAACAGCCAAC

TTGTACCTAGCTCTCCTTATTCTCCAAGGGCTGGGCCAGTTCTCCCCACA

GCCCTGCAAGGGAGGATCACTCAAGGGCCCCAACTGTCTGACAATACAGC

35   CACACTCTGATCAGCCACCTGGGCATAGGCTCCATGCCATTGTCCTCCGC

CAAGACCTCAGACTGAAATGTTGGCTCCTCCCATGAAGAACCTGGGGCCA

AAGGACCAGAGTCCAGGTCCGTGGCTGCCAGGATGGGCCACTTGGAGAGA

GGCACAAGGGTGGTGCCAGGCAGGTGTGAGGGCTGGACCTTTGCAAGAGC

AGCATCACTTTTGTTGAGAGCCCACAGGTATCTTATAATTGGGTCCTAGG

5   ACTTCCTGCCAGTAGCCATTGTGTGCATGGATTTGGGTGCTGGCCTCACC

ATGGTGTGCTGGCTGCCCATGCCTGCAATAATGACTTCTGTAAGCCTTTC

TTCATCTGCAAGATGGGTGCTGCTGGCACCTCCTCCCCGGTGCTGTGGTG

ACAGGGCATAGTGTGTGAGGCTGCTATGTGAAGCACCTAATGCAGGGCCT

GGCATATGGAGGAATTCAGCAAATGACAGATGCCTTCACAGTTAGTTCCT

10   GGCATCCTCTACATTGGTGGGTGTAGGAAAGAAAGACAGAGGAGGCAAAA

GTTGTAGCTGTGGGGCATTGAGGACAGCCTGGATTGTTCCACAGAGCCCT

GAGGACATCTCCAGGGGTGTGCTCTGCAGGGGCAGCTGGATTGGAGGGTT

AGGGGTCGGGGAGGGCGTGCACTCCCACCCATGCTCACAGCCTCGGAACA

GTGCCTGCTCAGCCAACATGGGTGTTTGATTCTGTGTCTTTTGTCACAGA

15   CTTTATCAGCCCCATCCCTTTCTGACCTTGCCTCAGTTTAAATTTTACAT

GTGGGGCCTCATTAAGAGACATGGTTCTTAACTAAAGATCTGTATCCATT

AGGAATGCTTTGGGCTGCAGGAAGACAAACACCTGACTCACTGTGGCATA

AGTGGTTTGCGTCTGCTCCCATAAGCTGCACGTGGAGGGTGGATCTGGCA

TTACTCTCTCTTCCCTACATTTGCAGTATGCTAACAGCTTTAACCTCCAG

20   CCTTGTTTCTTCATGGTTGCAGGGTGGCTATCACAGCGCTGGCCATCACA

TCCTTACACAGCTGTGTTTACAAATTTAGGGGGACATTGAAGCTCCTCCC

CTGCTAAAATCAGGCTTCCCTTCACCTGTCATTGGCCAGAACTGGGTGAA

ATGCCCAACTCTAGACCGATCATCAGTAAGAGGAGTATAGAATTGCTGTG

CCCACCTTAGATTAATCATGGCGCAATGTGCTCCCCATACCAACAAAATC

25   TGAGTTCTAGAAACTGAGGAAGAAGAGGAAAATGGCCGTCTTGCCTCCTG

GCTGGGATTCAGAGCATCTCCAACCCTCTGAGCTTATGTGTAAGACTGTG

GGCAAAGTGTGTGAGTTTTTGTGGAATGGATCCACGGCTTTTATCAGAG

CATCTTTCCTTTTTCTTTTTGATTCAAGATGAAAATATTCTTATGATTAT

TTTTCTCACCACTGCCCAGAGATAACCAGCACATTAACATGGCCTTTTCT

30   CCATGAATAGCACTAGGGTGCCCAGTGGACAGACACATAGCTGTCCACAC

ACCAGCTTGCTGGGGATGCATAGGCAGAGTCACATCTGCACTCACGGCCT

GTCCTCACACTGCCATGTGGAGAGCCAGCAGCCACACCATGGGCCGTCCA

TGCTCACGGGAGTGGCAGTATCAGATCTGAGCTTCGTGTGCCCAGGCGTC

TCTCACATCAGTGCATAGGGACCCTCTTTGTTCTGTGGCCCAGTGTGCCC

35   ATGCCACAGATGGCTTCAGTCAGCAGACACCTCCTTCTAGACACTCACAC

TCACTCCTGGCTGGCCCTTAGCACACCTGTGCAGACAGGCCCATTTATTT

TCTTGTGTAAATCCCAAGTAGGAGGACTGGGTCTCTCTGACAGCAATGCC

AGCTGCCTGGCACCCTCCAGACAGGTGGCTCAAGCCCCACCTCGCCAGCT

CTCCCAGTTAGCCCCTCCTTTCCCTGGCTCTGACCTGAGGGACGAAGCAG

5   GGTGCTACAGGACGCTGTGCCACAGGGATATCGTCAGGGACAGAAGCTAC

TCTGCCCTCTGCTGCTCACCCCTCCAACACGCTGTGGGCTGCATTTGTTG

AGTGGCTGGTACCAGACTCTGCTCTTCTGACTTTCCAGCTGGTTTTACCT

GTAGTAAAGTTTGAGAAGATGGGTCATCCTGACCCCGGGGTCAGAAGACA

GAAGGAGGCCCATGGCGTGTGGGGGAGATGCCCCGTGAGGCCCTCGGTGT

10   GCAGATGCCTGGTGACAGCCCCACCCTGAGGTCCCCAGCCTACCCCCTCC

CCAGCCCGACTGCTCCCATCCCCCTCCCTGTGCAGGTAGAGCAGATCCTG

GCAGAGTTCCAGCTGCAGGAGGAGGACCTGAAGAAGGTGATGAGACGGAT

GCAGAAGGAGATGGACCGCGGCCTGAGGCTGGAGACCCATGAAGAGGCCA

GTGTGAAGATGCTGCCCACCTACGTGCGCTCCACCCCAGAAGGCTCAGGT

15   ACCACATGGTAACCGGCTCCTCATCCAGAAGCAGCTGTGGGCTCAGCCCT

AGCTGGGAGAAGCACCCCAGGCACTCCCAGACTCACAGCCAGCCCGAGAC

AGAATCTCCTGGGGAGCAATGAAGTCCTCGACTTGGGCCAGTTCTCACCC

TTGGCTCCTCTGGTCCGGCCCTGGGGCACTCGGGCTCACCCTGGAGCTGG

CAAACCTCAGGAAAACTGGCGTTTTAAATCTCACTCCTGGCCAGGTGCAG

20   TGGCTCACCCCTGTAACTTCAACACTTTGGGAGGCCAAAGCAGGCGGATC

TCTTGAGGCCAGGAGTTTGAGACCAGCCTGCCCAACATGGTGAAACCCCG

TCTCTACTAAAAATACAAAAATTATCCAGGCATGGTGGCACATTCCTGTA

GTTCCAGCTACTCGGGAGGCTGAGGCATAAGAATTGCTTGAACCCGGGAG

GCCGAGGTTGCAGTGAGCCAAAATCGCGCCACTGCACTCCAGCCTGGGGT

25   GACAGGGTGAGACACCATCTCAAAAAAAAAAAAAAAAAAAGACCTCACTG

CTCCCCATGGGCACTTAGGGAACTCTCCCAGCCCAGTTCTGCAGCTGGGC

CATTGCACTAGATCCTCAGTTGGTCCCTGGGCTCTCGGTGACTGTCCAGG

GCAGGAGTTTCCCATTGACTTTTCCCTGGTTGACCTTTGACCCCTTCCAC

AGTTGACACTGGTGTCCCCAGGTGTCTGGTGGCCCCTTGTCCAGCTCCCT

30   TAGTCCCTTGTGCCTTCCCTCCTCCTCTTTGTAATATCCGGGCTCAGTCA

CCTGGGGCCCACCCAGCCCAAGGCCAGCCTGTGGGTGTCCCTGAGGCTGA

CACACTTCTCTCTGTGCCTTTAGAAGTCGGGGACTTCCTCTCCCTGGACC

TGGGTGGCACTAACTTCAGGGTGATGCTGGTGAAGGTGGGAGAAGGTGAG

GAGGGGCAGTGGAGCGTGAAGACCAAACACCAGATGTACTCCATCCCCGA

35   GGACGCCATGACCGGCACTGCTGAGATGGTGAGCAGCGCAGGGGCCGGGG

CAGGGGGCCAAGGCCATGCAGGATCTCAGGGCCCAGCTAGTCCTGACGGG

AGGTGCCACCTGTCTACCAGGGGTGGGGAGAGCGGGGGGCTGGAGGACCAC

CCAGCCTCAGAGGCAGCTGGAGGCCTGGGTGAACAGGACTGGCCAACATG

TCCCCAAGTCCCACAGTCACCATCTGGCCAGCATTGAGAGGGGAACGGGC

5   TGAGGAAGAGTTAGTGGCAAGAGGAACCCCAGCCAGTCACACCTTGTCCA

GTTTACCAGAGGAAAAACCAATGTGTAAGAACAGAAATGTGACCCGGCAG

CCAGTGCACTGCCCCCCTCTCCAAAGGCCACCCCTCACCCTCCACCAGCA

TGCACAGAAAGTGGGGTGACAGCAATCACAATGTCTACCCAGGCAGCAAG

GACCCCTGACCATGGGGAGGACTGGGGTGCAGGGAACATAGAAGCAGAAT

10   GAGGCCTAGGGGGAGTTGGGCAAGGCCAGAGCCCTAGCTGCAGCCAAGCA

CATGGCCAAGGCCAGCTCCTGGAAGGGCAGGGCTCCGAGGCAGGAGGCAG

GAGGCTGCCCGTGGCTACCCGTCCTCACACCCCTGCAGCTTGCTAGTCTG

TCTGTGGGCTGGGTGTGAATCAAGGCAGTGGGATGGTGTGGGGACCTCCC

TGGCCCCAGCAGCCAGTGAGGAGCCTGGTCAGTCAGCAGAGCATTCAGCA

15   GTATCCAGTTCCATGGAGAGGCCCGTGTGAGGGGAGTCGGGGCTGGTCTT

CAGTAAGGATGGGTGGCCAGGGCCCCTAGAAGTAGAAAAGGAGACTCCGG

GTGCTGGAGACAGAAATCAAGGATGTGCCTCCATGTGGAGCCTCAGGAAT

AGCTGGCCAGGCCTGAGGCTGAACCTCACAAGGTTCAGCTGGGAGGGCTA

GGCTGACAGAGCACAGCCGGGCCAGGGACCAGCCTGCCCTGTGTTGCCTT

20   GTCCCGAGGGCCACTGTCAGCAGGTCTCTGGCATGGGGGAGGCTTAGGGC

CTGAGCCCAACAAGCAGCAGCGGAAGAGGAGAGGGAAACTGTGGACAGGC

CTGGCATTCAGTGGCCAGGTGTTGCAGTGTCCCTGAGGAATAGCTTGGCT

TGAGGCCGTGGGGAGGGCTGCCGGCCAGCGCACCCCCCCATGCCAGATGG

TCACCATGGCGTGCATCTTCCAGCTCTTCGACTACATCTCTGAGTGCATC

25   TCCGACTTCCTGGACAAGCATCAGATGAAACACAAGAAGCTGCCCCTGGG

CTTCACCTTCTCCTTTCCTGTGAGGCACGAAGACATCGATAAGGTGGGCC

GGGTGGAGGGGCAGAAGGCAGATGAGGGGAGGCACAGGCACCCCAGAGGA

ACTCTGCCTTCAAATGTAGCCCCCATACCATGTGCTCAGAAGGGAGATCT

GGATTCAAATTGTGGCCATGTCACCTGCCACCTCTAATGCTGTGGAAAAG

30   AAGCATCACATTAGCTAATTCTGGCTGTGCGCCTTGTGAGGCACCAGCTA

TGATCACCCCACTCCAGTGGAAAGAGCAGCTGGCAGTAGGGTGGGGCTCA

AACTCAGGCAGCCGGGCTCTGGGTCACCTGCAGGCCACGGTCATGTCACA

CTGCCTCTAGCTGAGTCAGAAATGTGAAGGAACTGAGATTCTACCCTTCC

TGCAAGCTAGCAAAGTGGCCTGCCAGTTACATCTGTGCATGCACACACAC

35   ACACAGTTATATATGCACACACATAAAACACGAGACCTTTGGGTCAGGGA

38

GAAAGCCAGATCCTCACTCACGGCAGAAGCAGCAGCCAAAGCAACATCTC

ATGTGGTTTTCCAAGCCCCAGTCCCTACAGAGACAGAGAGGGCCAGGTGG

CACCTGTGCATGCAGCGGGGTACCTTGCAGGAGGGAAATCCTGATTTTAC

ACAAAGCTGCTCCCCCCACGCCCTGCCTTGACTCTGGGATGACGTCTCAG

5   AGCTGTGCAGTACAACATTCTTAAATTGGCTGGGACTCAGCCCTGCAGAA

ATATGATATCTTCAAGGAGAATCGTTCCCAAAACCTCTCAAAGCTATGGG

GCTGCTCTGAGCCTGTTTCCTCAGCTGTAAAGTAGGGTGCATACTTTTAT

GGCCCTGTGCAGGAGGTAGTGACAGGCCCTAGCACCCTGCCTCCAGTATA

TGTTAGCAGCCACGAGGCCTATCTCTCCCCACAGGGCATCCTTCTCAACT

10 GGACCAAGGGCTTCAAGGCCTCAGGAGCAGAAGGGAACAATGTCGTGGGG

CTTCTGCGAGACGCTATCAAACGGAGAGGGGTGAGGGGGCACCTGTACCT

GCCGGGGGGGCTGCCCTGGGCCACCCACCCCAGCACTGCCTGCCTTTCTC

CTTGGCTTCCAGCACTGCAGCTTCTGTGCTTCTTGGCAGGACTTTGAAAT

GGATGTGGTGGCAATGGTGAATGACACGGTGGCCACGATGATCTCCTGCT

15 ACTACGAAGACCATCAGTGCGAGGTCGGCATGATCGTGGGTAAGGGCTCC

TTGCACCCCTGCCCCTTCCAGACTGCTGAGGCTCCCTGTGTACAACAGGC

TTCAAGGGCCCTGTGGGGTGAGGACCAAACTACTTAACAACCGGTGATGT

CAGAGCAGAGCCTGGTGCTACAGCCTGGGTGGTCTTGGGGTATCAAGATG

GAAGCACCGTGTACAGTAGGAAGCATTTCAACGCCATGATGCCACATTCC

20 TGCATCAGATGGTATGCCAGCTGCATATCCACCTCACCCATCAGGATTAT

AATTAAAACACTTATCTGGTAAATTGACCAACTGGACAGATTGGTCCAAG

TGGAAGAGGATAAGCAAAAGTGGTACCATCTCCACCCGAATGGTCTTTCC

ACGGGCCTGCCCCTGCCCCTGCCCCCACCCAAAGTGAAGGCAGGTACCAG

GAAAGGGAGCAGCAGTCCGCCCCTCCCAGCAGAGGGGTCTTCCACACCAA

25 CTCGGACCTTTCTCAGAAGTTCCGGAGGTCATTATAACCAGCCTTCACTG

AGGAGCAATCCAATCAGATCAGTTATCTGCTGTGCGCACAGCCGTGTGGT

TCTATACTTCTCTTACTTCCATTTTCACCTTTCAGAAGGAACGTTGTCTT

TAAATCCAGCATCTAAACGTGAGCCCCAGCCATCCCTGGCTGTGATCCCC

CCAGCCCTTTCCACCCTATCCTCTGGAACTGCCTGGGGCTCCCCAAGACA

30 CTTCCACATGAATTCCCACCAAGCCAAGCTGCAGCTGCTGGGCCCAGGCA

TAACCCCTCCTGGGGCAGAGGTGGCAAGGAGTGACCCACCACTCACATCT

GCCCCACATCCACTCTTGACTCTGCTCAGTGTTTAAAAACATGTTTATAA

CAATTACCAAGATCTGAAAATTAGGAGAATTCACATCAAAGTCTGGATTT

CTGTTTGTTCATAAAAAACTAGAAGGCAGCCAGGCAAGGTGGCTCACGCC

35 AGTAATCCCAACACTTTGGGAGGCTAAGGCAGGCGGGTCACTTGAGGTCA

39

GGATTTGAAGACTAGCTGGCCAACAAGGTGTAACCTCGTCTCTACTAAAA

ATACAAAAATTAGCTGGGTGTGATGGCGCATGCCTGTAATCCCAGGTACT

CAGGAGACTGAGGCAGGAGAATTGCTTAAACCCTGGAGGCAGAGGTTGCA

GTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGTGATGGAGTGAGTG

5   AGACTCTGTCTCCAAATAAATAAATAAATAAATAAAAACTGGAAGTCTAA

GCATCACTGAGCCCTGATTCCTATGTGGCAGCTCGACTGACCAGCATTTG

AGTTGCTGTCCCTGACAGCTTTGGGGGTGTGCAGCCCACACAGTCATGCT

AGCTTGAGGCTCTGCTGTCAGCAGTTTGAAACTCTTAATAACTTGTGAAC

AAAAGACTCCATGTTGTCACTCTGCACAGGGGCCAGCAAATTACAAAATT

10   CCATATCCGGAATTGTCTACAGGAGCCTCTGGGCTGCTCCCAAGGGCCCA

CACCATGCCTTACTCACTTTGGGTTGCCATCCAAACATGTCTCATGACAA

AGAAGCTCAAACATGTGCATGGACAGTGCCAGAAAACAAGGGTCGTACAT

AGACAAAATAAAATGATAACGTCCCACAACCATTTCTTTGATACACACTG

TTTCTCTCAGTCCTCCCAACCACCTAGGTAACAGGCAGGGAAGGTGTTAC

15   TGTTGCCTGTTAGGAAAGAGGACAGCCCTGAAAGCTGTCCCTGGCCACTG

AAGCAACCCAGGTCTTCCAGCCCCAGGGAGAGCCGCCTTTCCATTGTTCC

AGACAAAGCAGAGACAGGCATGGGGGAGCGGGAGAGGGACTCCTGTGGGC

AGGAACCAGGCCCTACTCCGGGGCAGTGCAGCTCTCGCTGACAGTCCCCC

CGACCTCCACCCCAGGCACGGGCTGCAATGCCTGCTACATGGAGGAGATG

20   CAGAATGTGGAGCTGGTGGAGGGGGACGAGGGCCGCATGTGCGTCAATAC

CGAGTGGGGCGCCTTCGGGGACTCCGGCGAGCTGGACGAGTTCCTGCTGG

AGTATGACCGCCTGGTGGACGAGAGCTCTGCAAACCCCGGTCAGCAGCTG

TAAGGATGCCCCCCTCCCCCACAACCCAGGCCCTGGGCCGCTCTGGTGCA

GCGGCAGATGGGAGCCGGGCCATTGCAGATAATGGGCTTGTTTTTAAACA

25   ACTCTGGGGAAAAGCAAACTGACAATCCGTTCGTAAGCTCCATCCCTTCT

GCTCAGTCATGACCTGCCCCTGTGAGAGATGAAGGGTTAGTCCCAGTTGT

GATGTGATAAGCCCAGACCTCTTTCCTTCCGACAGGTGATCGTGCATGCA

GAGGAGGCTCTGAGACGCCCCCAGCAAGGTTCCTGGGTTTAACCCAACAT

TCCCCAAAGTATGTATTTGGCCACATTCACAGAAAGAATATTAGTCTTTT

30   GTGGAATGCTGCGGGTTGACAGTCACAGCTTGGAAACCAACCCACAGAGA

GCTCATCATTAATCATGGCTATCACTTGTTTACCACCTACTGTGCCAGGC

CTATGCTAATTACTTTATTAGCGTCCTCTCTGCCGCTCGCAGGCCTCTAT

TATTATAGGTCAGTAGTATTCGATTTATTTAAATTAAATACGGAAGGTCA

TAGATTAAGCAAGAAAGTGCCAGCAACATGGTGCGTGCCTCTGACTGGGC

35   ACTAACCCTCCAAGTCTTAGTTTTCCCAACCATAACTGGCCAATGAACAG

CAGCTCTGGATGCAGCTAAAGGAAGACTGAAGCTGTAGGTCCCGTGCTCG
GCGCAGGGCCCCCTGCAAGGAAGGTTTCGGAGGGACTGGATGGGGTCTTT
GAACTATCTGTCTTTCCCTTTACTGCAGTGGGCCCAGGGGCAGGCCAAAG
TTGCTCCCGTGATTGACTTGAACGTGCACGTTCCTAATCCCTGACATTTC
5   TAAAGCTCTGGCTCATTAACGAGGGAAAGACGTGAACCAGCTGGGGGAGT
GGGGATCGCAGTGCCCCACGTGGCCGCCTCGTGACCTCAGTGGGGAGCAG
TGGGGCCGGCTCCCGGCTTCCACCTGCATGAGGGGCCCTCCCTCGTGCCT
GCTGATGTAATGGACCTGCCCTATGTCCAGGTATGAGAAGCTCATAGGTG
GCAAGTACATGGGCGAGCTGGTGCGGCTTGTGCTGCTCAGGCTCGTGGAC
10   GAAAACCTGCTCTTCCACGGGGAGGCCTCCGAGCAGCTGCGCACACGCGG
AGCCTTCGAGACGCGCTTCGTGTCGCAGGTGGAGAGGTGTGCGGAGGAGG
AGGGTGGGTGCAAAGGGCAGGGGCTGGGGACGCCCGGGCACTGCAGACTT
GGTCTCAGGCGACGCTGAGTCCCAGGCCCGGGGCGCAGGGATGGGAAAC
TAGGGCCTGGGGCGGGATTCCGGGCGTGGGCGGGGCCCGGGGCGGGGCAC
15   AGGGGGCGGGGGAGTGGGCGGGGCCCGAGGCCGGGCGCTGGAGGCGAGGG
CGGGGCAGGGACGGGTCCAAGGGCAGGAGGCTGGGACAGGACGGGGATGC
AAAGGGAGGGGCGGGGCCCGAGACGGGGAGGAGGGGGAGGGCCCAAGGGG
AGGAGGCGGGGTCCGGACGGGGATGCCAAGAGCAGGGATGGGAGCGAGCC
TGCGTCCGGGCACTGGTCCCCATCCGTGAGTCCCCTCGGTGCTCCCTGCC
20   CGCCGTGGCCATCCTCTCACATCACTCACAACCCCAAGGCGCGGCATGGT
TGACACCCCCACGTTAGGACGGAGACCCTGGGCTTAGTTAGAGGGGGCAG
TACTAACCAGTCCCTGGCGGAAACGCTTTGGCTGGGTGAGGTGAGCGGGA
TCGCCCCCATTTCTCCAGAGAGGGGTCCCGGCTCAGCGAGGGAAAGAGGC
CGCCGCTGGGGGGACGGCTGGCCGGGGCCCCTCCCTGGAGAACGAGAGGC
25   CGCCGCTGGAGGGGGATGGACTGTCGGAGCGACACTCAGCGACCGCCCTA
CCTCCTCCCGCCCCGCAGCGACACGGGCGACCGCAAGCAGATCTACAACA
TCCTGAGCACGCTGGGGCTGCGACCCTCGACCACCGACTGCGACATCGTG
CGCCGCGCCTGCGAGAGCGTGTCTACGCGCGCTGCGCACATGTGCTCGGC
GGGGCTGGCGGGCGTCATCAACCGCATGCGCGAGAGCCGCAGCGAGGACG
30   TAATGCGCATCACTGTGGGCGTGGATGGCTCCGTGTACAAGCTGCACCCC
AGGTGAGCCCGCCCCGCTCTCTCCCTGGTAAAGTGGGGCCCAAAAAGCGC
GCGCTCCAAGGTTCCTTGCGGTTCCCAAGCTCCAAGATTTCGTAGTCCTC
TTCTCGTCCCCCTTGGCCTAGATTTGGGGGAAGGGTCGACTGCGTGCAGG
GCGCCCGGTAATGAATGTGGAGGATGAGGTGGGAGGAGGGACGGCAGCCC
35   TGCTTCTCTTCTGCCCAGCTTCAAGGAGCGGTTCCATGCCAGCGTGCGCA

41

GGCTGACGCCCAGCTGCGAGATCACCTTCATCGAGTCGGAGGAGGGCAGT
GGCCGGGGCGCGGCCCTGGTCTCGGCGGTGGCCTGTAAGAAGGCCTGTAT
GCTGGGCCAGTGAGAGCAGTGGCCGCAAGCGCAGGGAGGATGCCACAGCC
CCACAGCACCCAGGCTCCATGGGGAAGTGCTCCCCACACGTGCTCGCAGC

5   CTGGCGGGGCAGGAGGCCTGGCCTTGTCAGGACCCAGGCCGCCTGCCATA
CCGCTGGGGAACAGAGCGGGCCTCTTCCCTCAGTTTTTCGGTGGGACAGC
CCCAGGGCCCTAACGGGGGTGCGGCAGGAGCAGGAACAGAGACTCTGGAA
GCCCCCCACCTTTCTCGCTGGAATCAATTTCCCAGAAGGGAGTTGCTCAC
TCAGGACTTTGATGCATTTCCACACTGTCAGAGCTGTTGGCCTCGCCTGG

10   GCCCAGGCTCTGGGAAGGGGTGCCCTCTGGATCCTGCTGTGGCCTCACTT
CCCTGGGAACTCATCCTGTGTGGGGAGGCAGCTCCAACAGCTTGACCAGA
CCTAGACCTGGGCCAAAAGGGCAGCCAGGGGCTGCTCATCACCCAGTCCT
GGCCATTTTCTTGCCTGAGGCTCAAGAGGCCCAGGGAGCAATGGGAGGGG
GCTCCATGGAGGAGGTGTCCCAAGCTTTGAATACCCCCAGAGACCTTTTC

15   TCTCCCATACCATCACTGAGTGGCTTGTGATTCTGGGATGGACCCTCGCA
GCAGGTGCAAGAGACAGAGCCCCCAAGCCTCTGCCCCAAGGGGCCCACAA
AGGGGAGAAGGGCCAGCCCTACATCTTCAGCTCCCATAGCGCTGGCTCAG
GAAGAAACCCCAAGCAGCATTCAGCACACCCCAAGGGACAACCCCATCAT
ATGACATGCCACCCTCTCCATGCCCAACCTAAGATTGTGTGGGTTTTTA

20   ATTAAAAATGTTAAAAGTTTTAAACATGGCCTGTCCACTGTTCTTTGACT
TCTGTGCATTAGGACTGTGGGGACAATCTATAAAGAGTCTGCGTCACATG
CATGAAGACACTTCAGTATCTCGGCAATGCCCTCCAGACAGCTCCTCCAG
CCATCTGTGCCAAGGGGAGTGTGAGGAGTGACAGACCAGGCTGTAGGAAC
AGGAATGGGGTGTCATGGGGGATGGCAGAGCAGTGGACAGTACACTGCCT

25   GGCCCGGGCCCCTGCTTGCCTGCCCATGGAATGTGTGCAGAGGGAGTGCC
AGGCCAGGTGCTGCTCTGGAGAAGTGGGGGAATGAGGCTGGTCCTGCTGC
AGGTCAGTCTCAGCACCGTCCTGTCCAGTCAGAGTCACTTAGGTTTGCCA
GTGAGTAGGGGCCCAGATACATGTTGGATTTCTAAGGTCCCTCCAGATGC
TCCTGTCAGTGGAACGCCTATTTAGAGTTAGCCAAGCGTAGGCATAATGC

30   CATCTTTCTGCAGCATAAAATACAGTGACATAGAAACATATTTGTGTGAT
TTTCATGCATTCCTTTTTTGATGAGAGATATTACCCAGCTAATTAGGAAC
AACTGTTTTGTTTCCTTCAGATCATAACCCAAAGTTGTGATTTTGAAAAG
TCATGTCCCCCTTCAGATTTCTTGTTTTCTGCTACTTCTCATGTGGAATT
GCTTTGGCTCTTCTTAGTTCTCTTGAGTCTAAATTATTCCTTATAAGTTG

35   GTGCAAGCATCTGATTATTTTGTTATCATTACTGTTATGCTCAAGCATTC

42

ACAGAGTGGAACACATTTTAATATCAATTGCTTTCTATTTCTCCTTTATA

TTACAGTTCAGGACATTGTATTAATTATTAAAATTCTATTCGTAGGTAGG

TTATATGACTGAATTGAAATAGATAAAATGAATTTCTTTTCTAGATAACA

AAGGAGGTGTCATAAAACACTTGTTATGGGCCAGTGTGATGGCTCATGCC

5   TATAATCTCAGTGCTTTGAGAGGCTGAGGTGGAGGATTGCTTGAGGCCAG

GAATTTGAGACCAGCCTGGGGCAACATAGCAAGACCCCATCTCTTAAAAA

AAAAAGGGTGGGGCGGGGGGGCACTGCTGGGCGCGGTGGCTCATGCCTGT

AATCCCAGCACTTTGGGAAGCCAAAGCAGGTGGATCAAAAGGTCAGGAGT

TCGAGATCAGCCTGGCCAACATGGTGAAACCCCAACTCTACTAAAAATAC

10   AAAAATTAGCCGGGCATGATGGCGGGTGCTTATAATCCCAGCTACTCAGG

AGGCTGAGGCAGAAGAATTGCTTGAACCCAGGAGGCGGAGGTTGCAGTGA

GCAGAGATTGCACCACTGCACTCCAGCCTGGGCAACAGAGCGAAACTCTG

TCTCAAAAATGAATTAATTAATTAAAAAAAGAAAAAAAAAACACTGGGCA

GGGTGGTGTGCACCTGTAGTCCCAACTACTCCAGAGGCTGAGGCAGGAAG

15   GAGCACTTGAGCCCAGGAGGTTGTCTGCAGTGAGCTCTACTCATGCCACT

GCACTCCAGCCTGGGTGACAGAGCTCAGTGGCTTACACCTGTAATCCTAG

CACTTTGGGAGGCTGAAGCAGGCAGATCACCTAAGATCAGGAGTTCGAGA

CCGGCTGGCCAACATGATAAAACCCCGTCTTTACTAAAAATAAAATAAAA

TAAAAAATATATATAAAAATTAGCTGGGTGTGGTGGCACATGCCTATAAT

20   CCCAGCTGCTTGGGAGGCTGAGGAACAAGAATGGCTTGAACCCGGGAGGC

AGAGGTGGCAGTGAGCTGAGATCGCGCCACTGCACTCCAGCCTGTGCGAG

AGTGAGACTCTGTCTCAAAAAAAAAAAAGGGAATTTAAGAAATTTAAAAG

AAAACTCTTGTTATATAAAAGGGTATTGGGTCTGACAGATAAGAGCTCC

TGCACTCTACCAGCCAGCTACTGACAGACATAGGTCTGGCTCCAGTGGAG

25   GGGCAGCAGCCAGTGAGCCCAGCCTGGGGTGGCCCACTCCTGCTGCCTCC

AGGATGTCCCCTGTTTCCCCAGCCCCTCTGCTGTGCCCTCGGCCCCAGAA

GCTGGCGAGACTGCTTCTCTGGAACAGCATCACGCAGGCCTGCCCATCGG

CCCACTGTGCACCAGGCCTTCTGGGGATACAGATGTCAACCAGGTGGGGT

GCTCAGGAGGGGCACAGAAGCCAGGAATGACAAACACATCAGCCACCAGG

30   CAAATGGGAAATGTGCCCCAGAAGCTCCCTGCTGAGGATGTTAGGGAGAG

CATTCTGAAGTAGTGTGGTTGAGATGAGGCTTGAGGAAGGCAAGGCTCCA

AACAGCAGGGCAGACTGGGAGCAAGGTAGACTGCATGGGAGGGCAGCTGA

TGGAGCTCCTTAACCCTCTGGAATTGCCCCAAAGCCAAGCAAAGTGTTCT

TCTTGGGGTCACAGCTAGCTCAGGGATGCCTTCTGCCCCTTGGTCAGAGG

35   GGCAAAAGGTCAGAGCCTAGGGTCACCAAAACCTCTGGGAAGCCCCGGGG

GTCTCAGGCCACAGACCATCCTCAGAACTACACACTGCCCTCCCATGCCT
GGCGGGGGCCCTGGACTGGCCCTCACCAGCTGTCTTCTTGCACTGGCCAG
GGTTCTGGCTGGACTGGCAAGGAGGGGTGGTCAGATACAGGAGTAACTGG
ATCCCTTCATCAGGACCTAGGGTGGTGAGAGCTTTGAGCCTGCTCTGCTC
5   CAGGCAGACATTGTGTCTGGCCCTGCCAGGATGGATAGACAGCAGGATGT
TACACGTTGAGGACATGAAGGTCATCAGGAATGTGGCTGGAATCTGTTAG
GCCTCCCCCAGCCCAGGCGGGGGCTGCCAAGTTTGGGCCTATCCTCTGTT
CCTCTCCTTATTTGGACCTTCAGGTGATAAGGCTGAGACATAAAGGAGGC
TGGGCCCTGCCACCACGACAGCAGCCACACCTCTGCAGAGAATGGTGA
10 GTGCCTGCTGGGGAAGAAAGGCTAGCGGTCTCCCAGGTGCTGGCCTTTGG
GCTGGGGGAGCAGAGTTTTCTGTGCTTGTGTTGGGTTGAGGGTGGTCCCC
AGGGAGAGGAAGAGGATCCTGGCCCTGGCTCTCCTGGGAATGCTCTGGGA
CTGTGCATGATGGGTGGGGTGGGGAGACTCTGAGGAGTTGGGGAGAGGAC
CCCTCCCTACTCACAGTGTTGCAGGCCAGCAGGAAGGCGGGGACCCGGGG
15 CAAGGTGGCAGCCACCAAGCAGGCCCAACGTGGTTCTTCCAACGTCTTTT
CCATGTTTGAACAAGCCCAGATACAGGAGTTCAAAGAAGTGAGTGCCCAC
TCCCAGTAGCCTCAGATCCCATCCTGGCCCCCCCACCCCACCCCACATAC
ATACCCCCCTTCTACCCTGACCTTGCCTCTCACACCACCCAGGTCTCTCC
CCCACCTCCCACCTTCCCTAGAGCTGGGGGCTGCTCCCACCTGAAGGCCC
20 CCATCCCACAGGCCTTCAGCTGTATCGACCAGAATCGTGATGGCATCATC
TGCAAGGCAGACCTGAGGGAGACCTACTCCCAGCTGGGTGCGTGCACCCA
CCTCCCACCCTGCGCACTGGGGTCCCTACTCTGAGCTGCTGGGCGGGTGG
GAGTGGCTGGGGGGACAGGACTCTGCTCCCCTGCTTCCCCTCCTCCCCGT
CTCCTCACACTGCCCTTCCCCCCCTTGTCACGCCTTGCTTCCACTTCACCT
25 TCCCGACCCACAGCTGCCTCTGCCCCTCCAGCCCCTGTGGCCAGGATGGA
GGGAGGGCGGCCTGGGCCTTCTGGGGGACACCCAGGGTCCCTGTGTGCAC
CTCATGCCCCACCCCCACCAGGGAAGGTGAGTGTCCCAGAGGAGGAGCTG
GACGCCATGCTGCAAGAGGGCAAGGGCCCCATCAACTTCACCGTCTTCCT
CACGCTCTTTGGGGAGAAGCTCAATGGTGAGCCTGGGACAGAGCTGGGCA
30 CCCTTGGCCAGGCAGGGAGCCTGCACCCTGCCTGAACCCCACCTGAACCC
TGCCTGAACCCCACCTGAACCTTACATGAACCCCACCTGAACCCTAACTG
AACCCCACCTGGACCCACCTGGACTCTTCCTGGCCATGACCCATTCCAAG
CACATCCTCTGCCCCAGAATCCCATGTGCACTGGTCACCCCAGTGCTGAC
TTGGAGCCAGGAAATGTGCCTTCAGCCCCCACCCCCAAATTCCAGTCTCC
35 CAGCCAAGCTGCCCGCCTCAGGAGGATGACCATTCCCAGCCCCACTGATC

44

CCCGAGAAACATTTTATGTTAGGGAATACCCCCACCTCTTCTGGGATGTG

GGAGGCTCCTCATGCAGCCCAGTTCCTCCTGCGGGGGACCTGGGATGCTG

GAGACATGGATGCTCACCTGGCTGCCTCGGCCTTCCAGGGACAGACCCCG

AGGAAGCCATCCTGAGTGCCTTCCGCATGTTTGACCCCAGCGGCAAAGGG

5 GTGGTGAACAAGGATGAGTAAGTATGGGCCCAGCCAGATGAGGAGCACCG

TGGTGGAAGCAGAGAGCGGGGTGAGGCCCCTAGTGAGGGGGGCTGCCTGT

GCTTCGGGGCCTTACACTGCTCTTTGGGGTGCAGCCAACCCTTCCCTGCG

CCATGGGAGCCTCCGTACCCACCTTCCCTGTGCAGTCACTCCCCCGCAGT

CTCCTGCTCAGACCCTCCTCACCCCCCAGGTTCAAGCAGCTTCTCCTGAC

10 CCAGGCAGACAAGTTCTCTCCAGCTGAGGTGAGGCTGCCCAGCCCCTTCA

ATACTCATCCCCAGCACCTTCTCTGGGCCTTCACCCATGACCCAGAGCCC

AGTACCAGTGAGGCAGTTGCTGGAAGGGTGAGCCGAGGGCCCTTCTGGAG

GAGGTGCCATCTCTGTTGAGACCTAGAGGGTAAAGATGTGGAGTCAGAAA

AGAGGGCAGGGTGCGCCAGGCAGGGAGACTGTGCACAGACCTGGGGGGAA

15 GTGGATAGGGAGAGGTTTCGTACACTCGGGGTGGGCCTGTGCCTGTGGCT

GGAGGGGCGTCCTTTGCCTCTTGGCCCACATTTGCACTGACTCCTCACTC

TGCCCAGAGTCAGCCAAGAGAAAAACATTAACCCAGAGTCTGGGGTCTAG

GGTTGAAAAGCTAAGGCAAAAGCACAGATGCAGGGGGCAGACAGAAAGG

CCACAGGACTCAGGTGAGGTCTCTGCCGGGCTGGGCCAGGAGCCAGGGGA

20 CTGCCACTCACCAGTGTCCCCTGCAGGTGGAGCAGATGTTCGCCCTGACA

CCCATGGACCTGGCGGGGAACATCGACTACAAGTCACTGTGCTACATCAT

CACCCATGGAGACGAGAAGAGGAATGAGGGGCAGGGCCAGGCCCACGGG

GGGGCACCTCAATAAACTCTGTTGCAAAATTGGAATTGCTGTGGTGTCTT

GTCTGTGACAGATGGGTTGGGGACCAGCCAAGGGGGATCCCAGGGTCTCA

25 GTGCGCACATCACCATGATCATGGCCACCATCTACCTCCTGGGAGCTGGC

CCCTCGCCAGCTCACCTTGATTCACTCCCATGATGCCAAGTGAAGTGTGA

ACTATGATCATGCCTAGTTTACAGATGAGGACACTGAGGCCCAGAAAGTG

TGAGCATCTTACCAAGGCCAGCCCTCTAGAAGAGGAGATGGTGGGATTTA

CACCACCTCCACCAAGCCCAGGAATGAGCCACAAAGTGGGCACTGCCCAG

30 CTACTTGGGGCTGTGCAGAGAAGAGGCTGCTTGCTGGGCACTCAGCAAAC

TCTGCCCAACAGCCCAGCGGGTGGGCAGCAGCCCTGGGACCCCCACACCC

AACCACACAGCCTCCCCTGGCCCACTGCTCGCACCCCATCTCAATACACT

GGCTTGGGTGCCTCCCTGCATGGGCCCTTTGTGAAAGGCAGAGAGGTACC

CATTTGAAACACAACCAGCTTCTCATTGCAAATACAGGCAAGGCACTAAG

35 ACATGAGGAACATGGACACCAAAGCAGGGGCCAGGTAACATGCAAATTTC

45

TAGAGGAAATGCCCAGAACCTGGCATCATGCCTCCTGAGCCCCTCATGCG

CCGTGAGGGGTAAGAGGGTCAGACAGCTGGAGTGTAGGGAGACGACTTCT

CAGGAGAGAATAGTTAGTGCTCCCGTCACCCTTCATCTGAGAACCCAAGA

GCTAGAGGAGAAAGTGATCCTCATGAGTACCAGAGGAGCAGCAGGGGACA

5 TCCAAAGCACCAGAGAGAGAAACAGAGACAGAGAGACAGGCAGTGACAGC

TCAAACCTCAGCCAGATCCAGAGCATACAAAGTCTCCTGCCTACAGGACA

GCCCAGTAAGAGCTCTCAGCTTGCCTCCTTCCCTCCCCACAAGCCCTGCT

GCAATCCCTGTACCTGGGGGTCAGTGGGAAGGAGGTGAGCGAGAAAGGAG

GGGCACCCCTTCCTGAAGGCCCCAAGAGGAAAGGCGTTTTCACCCAGACA

10 GGTGTTCAGTTTTGATTTTATCTGGCGCCTGGCAATTTAATTACTAAATT

GAAACTTGAGACTTTCTGGAATTATGGCATTTTCTGTTGCTTAGAGAGAT

TACAAAAGTCACGAACTGCCTGAGTTTCCATCCTGAAAGCAGGCCACCAG

CCCACTCCACTGACCATGCTGGAACAGTGGATGAACAAAATCAAGTACCA

TTAGGATTCTACCACATGAGTCTGCTTGTTCAACAAGCTGATTTCATAAA

15 GTAAGGGATCATGTTATAATCCAAGCTCTACAGGGGTAAATTGTGAAAGA

CTAAAATGAACCAAAAAGATCATAGGTGTCCAGTTATCTGATTTGATGGG

GTGTCTGAACCTTTTGTTATCTTTGAGCTGTTTCAAAACTCTCTAAATTA

TTATTATTATTTTTGAGACAGAGTCTCTCTCTGTCACCCAGGCTGGAGTG

CAGTGGCATGATCTCAGCTCACTGCAACCTCCACCTCCCAGGTTCAAGTG

20 ATTCTCATGCCTCACCCTCCCAAGTAGCTAGTATTACAGATGGGCACACC

TTGCCTGGCTAATTTTTGTATTTTTAATAGAGACGTGGTTTCACCATGTT

AGCCAGGCTGGTCTCGAACTCCTGACCTCCGTTGATCCACCTGCCTCTGC

CTCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCGTGCCCTGCCACAAC

TCTAAATTATAACTAATAGCAAGGCAATGGTTCTTCTCTATTAACGTGCA

25 AATAAATGTTGTCCAGTGGAAGCACAACTGATTTTTCCCTTCTCTGTGGA

AGAAGCCAATTTTGCATCTATTAAGCAAATTCATCTGGGCATTCCTAACC

GTCTACACATGCACCGGCTCTTTGAATTCTTCTCTGAACCAGGCCCAGGA

ATAAGCCACAAGATGAGCACTGCCCAGCTCCTTGGGCTGTCACATCTTAT

TGATTCCCACATGAATTCACAAGTAAATAAAATATTTGGCGGTTGTTCAC

30 TTAGTATGCAAGTCAATATTTTGCTTTAAAAATATTATCCTTTCACACTC

CTGATATAGTTGTCTGATAAGGTTAGTCCTTCCCACACCAAAACTGCCTG

TATTAGTGTTGTTTGGAATAAACTGAGGGTAGAATGTATATGGTGTGTGT

ATGTGGTGTGTGTGTTTGTGTGTGTGTGTGTGTGAGAGAGAGAGAGAGAC

AAAAGAGAGAGACAGAAGGATAGAGAGAAACAGATGGGCACAGACCCAGG

35 ACATGAGTTCAGCCTACACTGACCAATATGACAGCCACTGGCCACTTGAA

46

ATGTGGTGTGAGTTGGGATATGCCAAAAGTGTAAAATGCACACAATATTT

TGAAGATTTCATACAAAAAGAATGCAAACATCTCATTAATAACTTTTAT

ATAGATCACATGTTGAAATGATAATGTTTTGGATATTAGATTATTACTAA

AATTAATTTCACCTATTTCTTTTCACTTTTTAAATGTGGCTACTAGAATA

5   TTTAGAATTCCATAAGTGGCTTGCATTTCTGGCTTTCACTCCTGTTGGAA

AGCACTGAGTTAGACTGTGTAGTACGTCTATTTAAGACTGCAGTTTCCAG

GCCGAACACCGTGGCTCACGCCTATAATCCCAGCACTTTGGGAGGCCGAG

GCGGGCAGATCACCTGAGGTCAGGAGTTTGAGATAAGCCTGGCTAACGTG

GTGAAACCCTGTCTCTACTAAAAATACAGAAATTAGCCAGGTGTGGTAGT

10   GCATGCCTGTAGTCCCAGCTACTAGGGAGGCTGAGGCAGGAGAATCTCTT

GAACCCAGAAGGGGAGGTTGCAGTGAGCCAAGATCAAGCCACTGCACTCC

AGCCTAGATGACAGAGCAAGACTCCATCTCAAAAAAAAAAAAGTAGAATA

AAAATAAATAAATAAATAAAGACTGCAGTTTCTGGGAGACTCTGAGGCAG

GCATTAGCCTTCTCTGCAGAGAGTACTTGCAGCAGGGAGCAGCAGTTTTG

15   ATGTCCTCAAAAGGAGCCAATTTCATTTGGGTAGGGTTGCCTCTGAGTAT

TCTAGCAGTACAGACAGAAAGGAGAGAAGGCTGTTTCCAGAAAGCAGAGA

TCATACGAATTACTTGTGAGACCAAACTTGTTCCTCAGGTGAAGCTCAGG

CATCCCTTATGTGGAGTGTCTAACAGTCTACACCTGAGGATGTTGGACAT

AAGGGGGTGTGAGGTGGGCATGGCTGGGGAGAGCTCTGGGAGGGGGAAAA

20   CCAGCTCCATGTTGTCCACCCACTGAAAGGAAAGCTCCCTCTGGGGGAGG

TAGATGCCCCCTGGCCAGGCCTGCAGGGCCCTGCTCACTGTGAGCCCTGT

GTGGTCCTGGCCTGGGTCCCACCAGCCATTGCCAGGCAACAGCTCCCAGT

TGGAAAACAGAGCAAGGCTCCCTCTTAGAAAAAAAAAAAAGAAAGAAAGA

AAAGAAAGAAATACAACAGGTAACTAAGCATGACGGCTCACGCCTGAAA

25   TCCCAGCTACTTGGGAGGCCAAGGCAGAGGATTGCTTGAGACTGGGAGGT

TGAGGCAGCAGTGAGCCAGGATTCTGCAATTGCACTCCAGCCTGGGTGAC

AAAGTGAGACCCTAGTAAAAAAAAAAAAAATAGAGACAGAGAAAGAAAGA

CATGCAACAGGGCCAGGCGCAGTGACTCATACCTGTGATCCCAACACTTT

GGGAGGCAGAGAAGGGAGGATTGCTTAAGACCAGGAGTGCAAGACCAACC

30   TGGGCAACATGGCAAAAACCCATCTCTTCAAAAAATAAAAAAATTAGCCT

GTTGTGGTGGTGCGCACCTATAGTCCCAGATATTCAGGGAGCTTGAACCA

GGTCCAGGCTGCAGTAAGCCATGATCGTGCCACTGCACTCCAGCCTGGGT

GACAGAGCGAGACCTTGTGAGAAAGAAAGAAAGAAGGGAAGGAAGGAAG

GAGGGAAGGAGGGAAGGAGGGAGGAAGGGAGGAAGGAAGAATATAGGACC

35   CAAAGGCCTAAATGCCCCTACTGTGCCCCAGTTCTGCGTGACTCAGGACC

47

AGCCTCCTCCACACTCCCACCACCACAACCCTGCACCCTACTTGTTCCTG

GGGGCCCCAAGGGGGAGCCTCACCAGAAGCCTCCTCATAAACCCACTGCCC

CTTACCTTTCCTGTCTTTCTAGAAGCCTCAGAAGCCTTGCCACTCTAAGG

ACACCTCCATCTGAGCCAAGGCGCTCGCTCCAGATGTCCCAGAGCTCCTG

GTCCTGGGTGTCCCTGCCACACAACCCCCCATGGAGCCCTGCTCTGGCTC

AAGCCCCCTGACTGTGCATGAGCAGGCCTGTTGCCCTCACTGGGACTGTC

CAGAGCCTTCCCATCTCTCTGGAGGGACTTCCATCAGTTTCTGCCCCTTC

TCCTCTGCCAAGAACTCACGTTCAGTCTGATAGCAGAAGAATCATCTGGC

ACCCTCCTGAATGGAACCCAGAGTACCTCCTTTGTGGACCGGTCTCTGGA

TTTTCCCCACTCTCTCCCTTCAGCCATGCTGATGGCAGAGAAGGTAAGAA

CTTCCAGCCCACTTCTCTGGCGAGGGGAACTTGTCATCTGGGTCTGCAGA

GAAGGTTCCACCTTATGCTCATAGTACATTATCTTTACTATGTACTAGGA

TATCACATTTAAAAGGACAAAAAAGGCCAGGCAGTGGCTCATGCTTGTAA

TCCTAGCACTTTGGGAGGCTGAGGCAGGTGGATTACCTGAGGCCAGGAGT

TCAAGACCAGCCTGACCAACATGGCGAAACCCCATCTCTATTAAAAATAC

AAAAATTAGCTGGGTGTCGTGGCATGTGCCTACAATCCCAACTACTTGGG

AGGCTGAAGCAAGAGAATCACTTGAACCCAGGAGGCAGAGGATGCAGTGA

GCTGAGATCGTGCCACTGCACACCAGCCTGGGCGACAAACCGAGACTCCA

TCTCAAAAAATAATAATAATAAATACAACAAAATAAAGAACAAAAAAA

AAGAAATGTAAAATACTTGAAGGGGCTTGTATAACATTAATAGGATTGAC

AGTATCTGCTTTCCAGGCTGAAGTGATTCATTCATTATTCTAGACGTCTT

TAGTCCTTTGCAATTTGTGGTAATTAGGCTTTTCTTTTTAACATTAAAAA

TATACAAAAATAAAAGGCAAAAAAAGCATCATCCCATTAGTCTGACCTTC

CCCTCCTCCATCCCTGCCCCAACACCCTGAAGACCCTGGATGCAAACAAA

GGCCCGAGGGAGCCTCTTCCCTCGCAGTGCAGGCCTCACCTGGGGCTCAG

AGTCAGAATCTGCATTTTATTCCCTAGGACAACCTCTAGTCAGGGCAGAG

GCCGGCTGTGCTGCCCAAGTGCCCTAACCCTAGCTTTGAGGCACCAGAAG

GGCAAATGCAAATTAAAAATGAGAATAAGTTTATTCTCCTTGGTGAAAAA

AAAAAAAAAAGACTTTCCCCTCTCCTTTTTCTTTAGAAAATCTATCATTG

CAAGTTCCTTCCTGGACTTTTTTTATGTAGATCTGTTCAAAAGCTAAATA

AGCCTCTTTCAAGTTTCACATCCCAGGAATGTCTCCTTAAGGACCTAGGA

GCCACCATTTGAAGTGTAATCACCAAGGGAGATACATCCTTATCTCCCAG

TTTCCGTGGGCAAAGGGGGAGCCTAACTTTAGCCCGGTGCCTAGCTCAAGT

TGCAAACACACTTCCAGTCTTAAAGGAATGAATTTATTTTTTTTCCTTTA

GGCAAACCCAGGTAGCCACCACAGTTACCTGGGGATTCACAGAGAACTGT

GTGTGACCACTGGTGCTGTCAAGTCCTCTTACCTGAGCACCTGTGACGTT

TCCCTTGAGAACGTGTACGGGATGGGTTGCACCTGGTTATATACAAGCGT

GAGACTTCTTTCTGCCTTTGTAATTTATTAGCAGATTATCTGTGATGAGC

ATCGCAATCTGTTTAATGCCTATTCAATAATTAAATTTTTCTTTCTCTTC

5   TTTTGTGGAAAGGTTTTCTGCATTGGCAGGAGATTTTTGTTTTCGATTAT

GTCCCCAACATGCCTGATGTTCCACCCCTCAAGAGCCTCAGCCTTGCCCA

GGGAGGGCATGGGGGTGAGTGGCCTCTCCCACAGAGAGTGCTGGCCAAGT

TGGCCCAGGTGCGCAGCAAGGGCTGCTGCCCAAAGGCTCCCTCCTGGTTG

10

The human liver glucokinase genomic DNA is 46,000 base pairs in length and contains ten exons (see
Table 2 below for location of exons).

15   The human adipocyte enhancer binding protein has the amino acid sequence depicted in SEQ ID
NO:3:

MAAVRGAPLLSCLLALLALCPGGRPQTVLTDDEIEEFLEGFLSELEPEPREDDVEAPPPPEPTPRVR

KAQAGGKPGKRPGTAAEVPPEKTKDKGKKGKKDKGPKVPKESLEGSPRPPKKGKEKPPKATKKP

KEKPPKATKKPKEEPPKATKKPKEKPPKATKKPSGKRPPILAPSETLEWPLPPPPSPGPEELPQEGG

20   APLSNNWQNPGEETHVEAQEHQPEPEEETEQPTLDYNDQIEREDYEDFEYIRRQKQPRPPPSRRRR

PERVWPEPPEEKAPAPAPEERIEPPVKPLLPPLPPDYGDGYVIPNYDDMDYYFGPPPPQKPDAERQT

DEEKEELKKPKKEDSSPKEETDKWAVEKGKDHKEPRKGEELEEEWTPTEKVKCPPIGMESHRIED

NQIRASSMLRHGLGAQRGRLNMQTGATEDDYYDGAWCAEDDARTQWIEVDTRRTTRFTGVITQ

GRDSSIHDDFVTTFFVGFSNDSQTWVMYTNGYEEMTFHGNVDKDTPVLSELPEPVVARFIRIYPLT

25   WNGSLCMRLEVLGCSVAPVYSYYAQNEVVATDDLDFRHHSYKDMRQLMKVVNEECPTITRTYS

LGKSSRGLKIYAMEISDNPGEHELGEPEFRYTAGIHGNEVLGRELLLLLMQYLCREYRDGNPRVRS

LVQDTRIHLVPSLNPDGYEVAAQMGSEFGNWALGLWTEEGFDIFEDFPDLNSVLWGAEERKWVP

YRVPNNNLPIPERYLSPDATVSTEVRAIIAWMEKNPFVLGANLNGGERLVSYPYDMARTPTQEQLL

AAAMAAARGEDEDEVSEAQETPDHAIFRWLAISFASAHLTLTEPYRGGCQAQDYTGGMGIVNGA

30   KWNPRTGTINDFSYLHTNCLELSFYLGCDKFPHESELPREWENNKEALLTFMEQVHRGIKGVVTD

EQGIPIANATISVSGINHGVKTASGGDYWRILNPGEYRVTAHAEGYTPSAKTCNVDYDIGATQCNF

ILARSNWKRIREIMAMNGNRPIPHIDPSRPMTPQQRRLQQRRLQHRLRLRAQMRLRRLNATTTLGP

HTVPPTLPPAPATTLSTTIEPWGLIPPTTAGWEESETETYTEVVTEFGTEVEPEFGTKVEPEFETQLEP

EFETQLEPEFEEEEEEKEEEIATGQAFPFTTVETYTVNFGDF

35   and is encoded by the genomic DNA sequence shown in SEQ ID NO:7:

CAGCAGGGCCAAGGTCTTGTGACAATGTCTGGAGGTGCCCCTATTGTCACACTGGGGGTCTCC

TACTGGCCTGCAATGGGAGGAGGGGCTGCAGCCCCACATCCTGTGCAGAGTGCTAGTGCTGA

GGCGGAACCCTCCTCAGAGCTGCCCCTTCTCCTCCAGGTTGTTACCCCTTCTACAAAACTGACC

5 CGTTCATCTTCCCAGAGTGCCCGCATGTCTACTTTTGTGGCAACACCCCCAGCTTTGGCTC

CAAAATCATCCGAGGTAATTTTTGTCTTCTGGGGGCCCAGGCTGATTTGCTGATTTGCTCTCAC

CTGGGGACAAGGTTCACAGAGAAGAAAACCTGCATTGTGGAGTCCCCCTGGCCCTTGTGGGA

TGGACAGCTGAGGTCTTCTGCACAGCTGCCATTTCACTGTGGGAGCCAAGCTGCCTCGCCAGC

TGGGCAGGGACTGGAACGGCTCCCAGCCTGTGTGCCTCTCAAGGCTAATCTCTGGTCTCCT

10 ATTGTCACTGCCCCACTGTGTGCCAATGGGGACTCCTGTTTATTTCTGGCAGCTTCTCTTTGAG

GCAGGACTTACTTGGAACCTACAGTGGGTCCTATGTGACTTCTTTGCAGGTCCTGAGGACCAG

ACAGTGCTGTTGGTGACTGTCCCTGACTTCAGTGCCACGCAGACCGCCTGCCTTGTGAACCTG

CGCAGCCTGGCCTGCCAGCCCATCAGCTTCTCGGGCTTCGGGGCAGAGGACGATGACCTG

GGAGGCCTGGGGCTGGGCCCCTGACTCAAAAAAGTGGTTTTGACCAGAGAGGCCCAGATGGA

15 GGCTGTTCATTCCCTGCAGTGTCGGCATTGTAAATAAAGCCTGAGCACTTGCTGATGCGAGCC

TTGAGCCCTGGGCACTCTGGCTATGGGACTCCTGCAGGGGTGCCCACAGTGACCATAGCCCAT

GCACCCACCAGCCGGTCTCCCTCCTCCCCATCCCTGACACCTCAGAATGTGAGCAGTCCGTGC

CATGAGCTTGTTTTATTGGAGTGACCTTGGCTCCCTCCCTCTGCCCCTACTCCAACACTGCAGC

AACCCCATCTCTTACGAGACTGGCAGGTGGAGCAGGAGCCTCTACACAGCCTCTGGCTCTTAG

20 GTCCCAGTCATGTTTGCACCCCCTCAAAGGGGCAGGACCAGCCCTTCCTTTCAGTGTCCATAC

CAGGGGCCTTCCATGTGCTGATGGGTGATGTGACTGTGGTCAGCAGGCTTGGGAAGTGC

TGCTGCTGTAGCTTGAGTTGGGCTGGGGTCTTGGTAGGACGCTGATCTCAGAAGTCCCCAAAG

TTCACTGTGTAGGTCTCTACTGTTGTGAAGGGGAATGCCTGGCCAGTGGCTATCTCCTCCTCTT

TCTCCTCCTCCTCCTCTTCCTCAAACTCGGGTTCCAGCTGGGTCTCGAACTCAGGCTCCAACTG

25 GGTCTCAAACTCGGGCTCCACCTTGGTCCCAAACTCGGGCTCCACCTCGGTCCCAAACT

CTGTCACCACCTCTGTGTAGGTCTCAGTCTCCGACTCCTCCCAGCCAGCGGTGGTTGGCGGTAT

GAGGCCCCAGGGCTCTATGGTAGTGCTCAGGGTGGTGGCAGGGGCAGGGGGCAGCGTGGGAG

GCACAGTGTGGGGGCCTAGGGTGGTGGTGGCGTTGAGGCGCCGCAGCCGCATCTGTGCCCGA

AGCCGCAGGCGGTGTTGTAGGCGTCGCTGCTGCAGGCGTCGCTGTTGGGGGGTCATAGGGCG

30 CGATGGGTCTATGTGTGGGATAGGCCGGTTCCCGTTCATGGCCATGATCTCCCGGATGCGCTT

CCAGTTGGAGCGAGCCAGGATGAAGTTGCACTGAGTGGCCCCGATGTCATAGTCAACATTGC

AGGTCTTGGCGCTCGGGGTGTAGCCCTCCGCGTGGGCTGTCACGCGGTACTCACCCGGGTTCA

AGATTCGCCAGTAATCACCACCACTGGCTGCGGAGGGAGAACGATCCGGCTGCCCCAGAGCG

CCCCTCCCAGGCCCCCACCCTCCCACTCAGTCCTGCCCCCAGCCCCGCCCTCCCCCTCTGAGTT

35 CCCGCCCCCAGCACCGCCCTCCCTCTCTGAATTTCGCCCCCAGGCTCCCCAGACTCTACCTGCT

50

CGCTGAGTTCCTCAAGCCCCCACCCTCTCTGGCGGGTCCTCCCTCAGAAAGATGGGGTAAAGG

TGTGCACACTAGGTACCTGTCTTCACGCCGTGATTAATGCCACTCACAGAGATGGTGGCGTTG

GCAATGGGGATGCCTTGCTCGTCCGTCACCACCCCCTTAATGCCGCGGTGCACCTAGGGAAGC

AGGTGAGGGCTGCTGGTCCTCAGGAAGGTCCAATGTGGTCCGCTGCTCCCTCCCGCCCATCCA

5   GGAGCCTGTGCAGCCTCCTCTCCCCAGGCATTGCCCTAGCCACCCCACCTGCTCCATGAAGGT

GAGCAGCGCCTCCTTGTTGTTCTCCCACTCGCGGGGCAGCTCACTCTCATGAGGGAACTTGTC

ACAGCCCAGGTAGAAGGAGAGCTCCAGGCAGTTGGTATGCAGGTAACTGAAGTCATTGATAG

CTGGCCGGGGACAGATACAGACCCAAAGTCAGCCCCTCTCCGGACCAGGCCCCGCCCACAGC

CCCTCCCAGGCTGACTCACTCCCGGTCCGGGGGTTCCACTTGGCCCCGTTGACGATGCCCATG

10   CCGCCGGTGTAGTCCTGGGCTTGGCAGCCTCCGCGGTAGGGCTCGGTCAAGGTGAGGTGTGCG

GAGGCGAAGGAGATGGCAAGCCACCGGAAGATGGCGTGGTCTGGAGTCTCCTGGGCCTCGGA

GACCTCGTCCTCATCCTCCCCCCGGGCTGCTGCCATGGCTGCGGCCAGCAGCTGCTCCTGGGT

AGGCGTGCGGGCCATATCGTAGGGGTAGGATACTAGCCGCTCGCCGCCGTTCAGATTTGCTCC

CAGCACGAAGGGGTTCTTCTCCATCCAGGCAATGATGGCCCGGACCTCCGTGGATACCTGGAG

15   TGGCCAGCACGTGTGAGGCCAGGGCTGCAGCTCCGGCCACTATCCCCAACCTAGCCCGATCAC

CCTCCATGAAGCTTCACACCAGTACTCGCACGATCCCCTGTCCCCCAACCCCCAGAGCCTCAG

CGTCTGGAGTTCAGGCACCGTCAGCCCCACCCCCAAGCCCAGAACACCAGGACCCCAGGGTC

CAGCTGCTCCCTCCTGCCCTTTCAGCCAGGCTGTAGCCTCACCGTGGCATCTGGCGAAAGGTA

GCGTTCAGGGATGGGCAAGTTATTGTTGGGGACCCGGTAGGGGACCCATTTCCTCTCCTCAGC

20   TCCCCAGAGCACAGAGTTGAGATCCGGGAAATCTTCAAAGATGTCAAAGCCCTCCTCAGTCCA

CAGTCCCAGCGCCCAGTTCCCAAACTCTGAGCCCTGTGGGGAGCCAGCAGGGTAGGCATCGG

CTACCCACACCCCCACAACCCCCAGCTGCCTGGACCCTGGCCAGCCTCACCCTTCAACCCACC

ATCTGCGCTGCCACCTCGTAGCCATCAGGGTTCAGTGAGGGCACCAGGTGGATGCGTGTGTCC

TGCACCAGGCTGCGCACGTGGGTTCCCATCGCGGTACTCTCGGCACAGGTACTGCATGAGC

25   AGCAGCAACAGCTCTCGGCCCAGCACCTCGTTGCCATGGATCCCAGCAGTGTAGCGGAACTCG

GGCTCCCCTGCAAGGGCGGGAGCCTCAGTGAGCACTCAGTCTCCCGAGGCCCAGGGCAGCTG

AGGAAGGACCCAGACCCACCTCATACCCGAGGGTCTGGGGGACAGCTGGGGCTCCTAGGGCC

CTGTAAGACAAGCCAGAATCCCCAGAGAGGCTCCGGAACAGGCGGGAGGCAGTGAGCTCTGC

ACATCAGCAGCAGAGGCCAGCTGCTGGCCCCCACAGACCCTCCCCCAGTTCATGCTCCCCAGG

30   GTTGTCTGAGATCTCCATGGCATAGATCTTGAGGCCTCGTGAGCTCTTGCCCAGGCTGTAAGT

GCGGGTGATGGTGGGGCACTCCTCGTTCACCACCTTCATGAGCTGGCGCAGAGGGGGAGGAC

GTGGAATCAATCATGCAATCCGTCCCCGCTGACCATGCCCCTTCCACTTCCAGGGCCTGCTCT

ATGGCGAGGGACGGGCATGACCCCTTCACGCAGCCCCAGGTACTGGCCTCCTTCCTAAGGTG

AGGGACAGCCAGCATCCCTGGAACCAGTAGGGACTGGGCCCAGTGACAGAAGCACCAGGCAC

35   ACACTCCCGTCAGCCACAGACAGGTCCCACCCCCAGCCCCAGGATATATGCTCCCAACCTGGC

GCATGTCCTTGTAGCTGTGGTGCCGGAAATCCAGGTCATCGGTGGCCACCACCTCATTCTGTG

CGTAGTAGCTGTAGACAGCTGCAAGGGAGGCGGGGTTGTCTTTAGCTGGGTGCCGGCTGGCCC

ACCCTAGCACCCCACCTCCACTCAGAGCCCCTGCCAGCCCTCCACACTCACGGGCCACAGAGC

ACCCCAGCACCTCCAGGCGCATGCACAGGCTGCCATTCCAGGTGAGTGGGTAGATGCGGATG

5 AAACGAGCCACCACCGGCTCTGGGAGCTCACTCAGCACGGGTGTGTCCTTGTCCACGTTCCCA

TGAAAGGTCTGGGGAGAGGCAGGCCTCAGAGCAGTACTGCCAGCCCCTCTGAGAGCCCACCC

CTCGCCCAGACAATGGGAGCAGAGCCAAGAGCCTGGGCATGGTGCCCACCATTTCCTCATAG

CCGTTGGTGTACATCACCCATGTCTGGCTGTCATTGCTGAAGCCCACGAAGAAGGTGGTCACA

AAATCGTCACTGTGGAGTGGACAGTGGTCAGAGCAAGGGTCTTCCCCCTCCCAGGCCCTCAGG

10 TGGCCTGAGCCTCCCTCTTCCGAGCCCCAAGAATTTAAGAGCTAGCAGGGTGGTGCTGCACG

GCCCAGGTGTTGAGCCTGGGTCCTATGCCCGTCACATAGCCATGGGCAGGTGATCTGTCCCTA

AACTCATGTGCTATCAGGACACAGGGGCTGACTGACCAGGCTGAGGAGTGGGGATGGGCAGG

GTGAGTCCCTCACTGATCTTTTTGGCCTTCTTTGGCTGGGCCAAAGAAGGGCCCACTGGAATCT

CCTTAATGGGACACAGAGCCATGCCTATGTAGCCACTCCCCTCTGCCAACTATCCATGAGC

15 CTGGCCACGCACTGGATGCTGGAGTCTCTGCCCTGGGTGATGACGCCTGTGAACCGGGTAGTC

CTCCTGGTGTCCACCTCTATCCACTGGGTCCTGGCATCGTCCTCGGCACACCACGCACCATCAT

AGTAGTCGTCCTCAGTGGCACCGGTCTGTCCAGGGGGCAGGGGAGGCTGAGCATGGGCGGAG

GAGTCCCTTATCCCAGTTGGGAGATGGGCCCATCCCAATGCCCACCTGCATGTTGAGCCGG

CCGCGCTGTGCCCCCAGGCCGTGGCGCAGCATGGAGGAGGCTCGGATCTGGTTGTCCTCAATA

20 CGGTGTGACTCCATCCCAATGGGGGGACACTCTGAGGACGCGTACCCCAGAATGGTGGCTCA

CTAGCTCCATCCTTCCCTCCACCAAACCCAGAACCAAGGAGCCCAGAGCCCACTCCCGGCACA

TCGGGGGCACAGTCAGAGGGCAGCTCTGGTCAGCTGGTGGCTCCCTGGTGCCCTGCACCAGC

CCACCTGGAATCGACTCAAAGCCAGGCCAGGAGCTGTTTCCAATCCCAGCCTGTGCTTCCCCT

CCCTGGGCCTCAGCTGCCCCATCTGGAGAACGGGCTGACCATGCCCAGCTCTCAGGGGACACA

25 CGTGAAATCACAGGTAGAGCTCCCCCAGGGCGCAGCCACAGATGTCATCCAGATGGGGACCG

TCTGCACAATGGCCCTGCAGGGATACCTGTGAAGGTACCTGAGGTCCTCACTCCCCACCAAGG

CCCCAGGTCCTCCCCCTACCACGCCCAGCCACTAGGGGCCCTGGGGAGCTGCCACCCTCCTGA

AGCAGGCCAGCCTGGGGTCCAGGGCTGGGGCAGCCAAGCGAGGCTATCCTGGGCTCCCGGGG

CCCCTCCCTTCTGGGTCCCAAGAATCTGAGTAGGAAAGGGTTCCGGGGACCTGGGTCCTGTTT

30 GTGACATTGGGCCAGTCACTTGTCCCAGCACCCCCATCCTGTGGCCCCCACCCTCACCCCCTTG

TGCCCCCCACTTACTGACTTTCTCCGTAGGCGTCCACTCCTCCTCCAACTCCTCGCCCTTTCGG

GGCTCTAGGGACAATGAAGGGAGGACATGGCACCAAGGGCCCGGGAGGCAATCAGGAGTCC

AGATGCTGCCCCACAGGGACCCAGGCCCCAAGCCCCAGCCACACACCTTTGTGGTCCTTGCCC

TTCTCCACTGCCCACTTGTCGGTCTCCTCCTTGGGGCTGCTGTCCTCCTTTTTGGGTTTCTCTGG

35 AAGGTGCAAGGTAGGAGGGGCCAGTCAGCCTGGCTCTGGGCTTTGAGGACCATGTGGGGTGG

52

ATCAGGCAGGCCCCAGGTGGCCTTCAGGGCAGGCCTGGTGTGGGAAGTCCTTGGTCCCACTCA
CTCAGCTCCTCCTTCTCTTCGTCCGTCTGGCGCTCAGCATCGGGCTTCTGGGGCGGAGGAGGCC
CAAAGTAATAGTCCACTATGGGGAGGGAGAGCCAGCTGAGGCTGCCCTGACCCTGCTGCGGG
GCCTCAGCTCCTGGGTCCACAGGAGCTCAGCAGGACAGGACCGCGCCAGAGGGGAGGAGGAC

5   GGGAGATGGGGGACAGCTGAGTTGGGAGAGGGTCTTGCAGGAGTCAGGAGCAGCCCGAGCTC
AGGGGCAGCTGAGCAAGACCCTGCTGAAGTCACCAGCCCGGCCTTCCAGGAGCATCTGGCCT
GGGGAAAGGACTCGAGGCCCAGGGCATGGGAAAGGCCTGGAGGGACAACTGGCACCTGTGC
CTGGGGTTGCGGGCTGGGGGGTGAGATGGGGAGACATTGGAGGCACTGATGGGGACCTGGGG
GCAGGGAAATGGCGATGCACGGGCTGCCACCCAGGAGGAAAGGGAACCTGAGGGCTCCAGG

10 GACGCAGGGGCATGAGCAACAGGGAGGCAAAAGCCCTCGGGCTCCCTGAAGAGAGTGGGGC
AGTGGCCACGAGCCAGCGGGAAGCCAGTTAGAGCACAGGACTGGGAGGGCTGGAACCCACA
TGGGTGACAGGGCAGAGTGTGTGCCTAGGGACACCCCTGTGGGGGTCACAGCCAAGCAGGAA
CCAGGGAAGCGGCCAAGGAAAGACCAGCCTGAGGGCAGAGGAGACAGGGCAGTGGCTGGGG
TGGGCACGCAGGGACAGCAGGGACAGCGAGGTAACCACGGGCACAGGTGGGGTTGCAAGGT

15 GGGTGAGTTGCCCCAGCTGGCTCCTGACCACACCCCAGCCCCGACCCCCACCTGCCTATGTCC
CTCAGACTCTGGGGTGCTGGGTACTCACTGTCATCGTAGTTGGGGATCACGTAACCATCACCA
TAGTCAGGGGGCAGCGGGGGCAGCAGAGGCTTCACAGGAGGCTCTGGGGAGGCGGGGAGGT
TAGGAGGGGGCCAGAGCGCCGTGGCCATGGCACCTCCTCTCCTGCCCCCCATCCTACCAATCC
TCTCCTCCGGGGCTGGGGCCGGGGCCTTCTCCTCAGGGGGCTCTGGCCAGACCCGCTCGGGCC

20 TCCTCCTTCTGCTTGGGGGTGGCCTGGGTTGCTTCTGGCGCCGAATGTACTCAACTGAGGGGG
AGGCTGGCTCAGAGTGGGGCCCAAGGCTGGGATGGGCCCATTGGCACATCCCCCAGGCCAGG
GGTCCGACCCAGGTGGGGCTGGCAGGACCCTACTCAAAGTCCTCATAGTCCTCCCTCTCGATC
TGGTCATTGTAGTCCAGTGTGGGTTGCTCGGTCTCCTCCTCCGGCTCTGAGGGGAAAGCGCTG
GTAGCTGCCTGACAACCCCACCCAGGCCTACTCTGGGGAAGCCCTCAGTCCAACCAGCCAGG

25 GCAGCTGGCCCCAAGGCCAGGCGGATGACGGCCACTCACCAGGCTGGTGCTCCTGTGCCTCCA
CATGGGTCTCCTCTCCTGGATTCTGCCAGTTATTTGAGAGGGGCGCCCCTGCAACACAGGAGT
TCCAGAAGCAGGTGGGCGGGAGGCCTGCTCTGACCACCTTGGGAGCCTCAGGCCACCAGCCA
CCCATAGAGCCCACACAGAGCCTGTGGACACCCTCCTGAGGCCGAGCTCACTCCAAGGAGGC
CTGAGCTCCTCTGGCCTTCAGCATCCTGCTGGCATCTCATGGGGCCAGAGAGCTGGGCCCACC

30 TTCTGGGGAACCTACTGTGCTGCTGGAGGCCCTACCACAAAGCTGTCCCCAGCGGGAGAAGG
CAGGAGGGAACTCCATGGGCTCAGAGCCCAGGGACATCTGGGCAGGGGCCTGAGGGACAGA
GGTCCCACCCAAAAGGCTGCCAAGCCCTCTCCCTACCCAAAAGAGGCTACAGCACTGAGGGA
GCCCACCAATCAAATTGTGAAATTTATAGCAAAGTGAGGTTCCCATCCAGTGGGGAGCTGA
AGGTCTATAGGAAGCAGGGCCCCAGAAACCTGCCTCCCACTCCCTGCCTCCACCCGAGCAGGC

35 AGTCAGAGCCCCATCACCCCAGAGGAGCCCGGCACAAACCTCCCTCCTGGGGTAGCTCCTCGG

53

GGCCAGGGCTGGGGGGTGGGGGCAGTGGCCACTCCAGGGTTTCTGAGGGAGCCAGAATGGGG

GGCCTCTTCCCTGACGGGGGCTTCTTGGTGGCCTTGGGTGGCTTCTCTTTGGGCTTCTTGGTGG

CCTTGGGTGGCTCCTCCTTGGGCTTCTTGGTGGCCTTAGGTGGCTTCTCCTTGGGCTTCTTGGT

GGCCTTGGGTGGCTTCTCCTTCCCCTTCTTGGGCGGCCTGGGGGACCCCTCCAAGGACTCCTTG

5   GGCACCTTGGGGCCTTTGTCTTTCTTGCCTTTCTTCCCTTTGTCTTTGGTCTTTTCCGGAGGCAC

TGTCCAAGATGCAGACTCGTGTCAAATGAACAGAGCCAGCTCTGTGCCCCCATGAGGCCCCTC

TCTAGATGCCCAGAACCTGGGCACAGGGACTCTTGTCAGTTCCCAGTGCGGATCAGCAAACTG

AGAGGTTAAGTCATTTGCCCAAGTGGCAAACTGGGATCCGGACCCAGATTTTCTGTCTGCAAG

TCTGGGGCTGTGACCACCAATCTCAACCTCTCTAAAGACTGAGCGTAGGGTTCCCAGTTCCCA

10   GGGGGAGGCCCTCATCCCCCCACCTGCCAAAACCTCAATAGGGGTTCCTTACTATCCACTCCT

CCACTATTCTGTTCTGGGCACAGAAGGGGCAGAGAGGTGACTGAGCCATCCAGGCCTGGAGG

AGCATCTGGTCATCCCTGCCAACTGCCATACAAAGGAAGGGACATGGGCCCAAGACCTTCCCC

TGGTCTCCTACGGGGCAAGAAAAGCTTCAAAGAAAAGGGACACTTGGTTGAGTATTGAAGCC

CAAAGAAGAGGAAGTGGTCTCCTTTCGAGAAGTAAGGGGTTTGGAATTGATTGGAAGGATAG

15   GGAGTCCTGGGGGGGTTCAGGGATCACACAGAGGACAGAAAAGACAGGTAGGGAGCTTGTGG

CTGCACACTCATTTCAGAGTCTGGGAGAGGGAGCAGGGACTGGTTGTGAGGATTCCCCATGG

GAATCCTCCCAGGACCCTAAGCAGGAGCTGCAAGTGCTGTTGAGAACCTGATGAGAGGTGGG

GAGCATGAGGGAAGTTTGGCAGAAACACAGGAAAGCTACCAAATGCAGACAGCCAGGGGAC

GCAGGGCTGCTAGAGCGGTGCCCCAGAGCCAGGAGAGCAAGCCTGGAAGGAGAGCCAGAGG

20   CAGGAGGGGCACAGGCAGCCCAGGGTGTGGGAAGCAGCCAGGAAAGATCTAGAGCTGGGGT

GGCAGGGGAGGGGCTGCTGACATCAGGAATGTTGGATGGTGCCTTGGAATCTCCTGGGAGAC

AGGGATCACAAGACCCTCTGCCACCTTCCAGAGGGCCACGATGAAAACAGCTAAGATTTACT

GACAACTGATTATGCAAGAGGCCGTGGGTTAAATGCTTCAGTGATGCATCACCTCATCTAATT

TCCTGTACTAATGTAGGACCACCCATTGCTCACCACCACCTGAAGCCCTGTGCTCACCACCAC

25   CTGAAACTCTCTCACCTACGTGAGACCTCCTGGAGTAGGAGGGCAAAGGCAGGAGGGAGGGA

CGACGTGAAGCTGTGCCACCAACAGGGAGAGTGGTCCCATTAGTATGGCAGGGGGTGACACA

GCACAGTCCCCTGTGGCTCAAGCCTAGTACCTGTCGCGTACTGGAGGAATGGGGATAAGCGA

CCCGTACAACCACAGCACCAACCCTAGAGCCACCGGCCCCCAAAAGCGGCCCTGCCGCCCGG

GTGCTGGATGTGCCTCCACGCCAGCGCTGACCTCGGCCTAGCACAGGGTCCCTCCAGGCATCT

30   GGGCTCGCGTGCGCATTAGTAAGCCAGCCATTCCTCCCCTAGCAGACTGGGGAGTGGCCAGAC

CCTACCGAATCCCCCTGTTCCCACCTGAGATGCCAGCCCCCCACACCCCCGCCCTGCCCTGGG

CTCTTACCTTCTGCGGCCGTCCCTGGCCGCTTCCCTGGCTTGCCCCCCGCCTGGGCTTTTCGGA

CCCGCGGGGTGGGCTCGGGAGGCGGCGGGGCCTCCACGTCGTCCTCCCGGGGCTCAGGTTCTA

GCTCTGACAGGAAGCCCTCGAGGAACTCCTCGATCTCGTCGTCGGTCAGCACCGTCTGCGGGC

35   GCCCTCCAGGGCACAGGGCCAGCAACGCCAGGAGGCAGCTGAGCAGGGGCGCCCCGCGCAC

54

GGCCGCCATGGCCGCGGCACGCGCGGGGGGGCTCCGGGGAGGGCGCGGGGGGGTCAGGGGCTCT

GGGTCTCTGGGAAAGGGCGGAGAGGGGATCGAGACGGGTGAGGGAATCCAGGAAGGGGCGG

GAGAGAGGATGGGGTGAGCGAGGGAATCCGGGAAAGGGAGGGAGAGTGGATTAGGGTGGGC

GAGGGGACCCGGGAAGGGGTGCTGGGGGGCTCCGAAGCCAGAGGGGCTCAGGGGTGGTCGG

GGCGCTCCGAGGTCTGGCGGCTAATAGGCGCTCCGGCCCCGCGTGGCGCACTCCCGCGCGGAT

AGCCGTCTCCAAAGCGCTGGCGGGGCCCGGGGCGGGGGCGCCGGGGCTTCCGGAGCCGGCTC

CCCACCCCCGGGGAGGAGGAGGAGGAAGAGAAGGAGGAGCCGAGAGTGGACGGAGGGGCTG

CGGGGGGGCGGGGGGCGGGGGGCGGGGGGCTAGGGGCGGGGCAGGCGGGCGGGCGCTGGCG

GCGAGCGTCCCAAGCCCGGAGACTTGCGCCTAGGACAGAGGGGCAGGGGGCGGGGCGACTG

GGAAGACAGAGGGCCTGAGGGAAGGAAAGGTGGTGGGGAGGGCCTGGGGTGCGGGTCTGAG

GGGGCCGACATCCCTCCTCCTTCTGCCCTAGGCACCCCCCTTAAGGCGGGACCCCGAGTCCAC

CGGGGCTCTGAGCCCTCCGCGGGTGACCAGGAACCCTGGACGGAAAGCCGTGGTGTCAGGCC

TCTGAGACCTCTCTCAATTCGGAGGGCCACAGAAAGGCCACCCCATCCTTCCCAGGCTCTGGA

GCCTCTGCCCATGGGCCCTGCTGCATCCCAGCGTCAATTCATTCAGTCATCCTACCAACCTCTT

CAGGTCGGTGTGGGGCCGGGCCCCGTGCTGGGCCCCAGGGAGGGACAGCACAGTGGGAACTC

ACTTTCCAGCCAGGAGGCAGGTGCAAAACTGCCCTCAGAGTGGCCAGCTGCCCCGCTGGGGG

TAGGAGTCCCATGTAAGGGCATGCCATCCCTCCCCTCCGGGTCCCAACGTGGACAAATAGCCA

TTTATCACCTTCTTCTTACCAGAACTCATTTTTTAAAAAGTGTCTACCATACCTCCAGCTGCCA

CATGGACCCAGAGGGCCCAGAGGACCCAGAAGGCAGGTGGATTGAGTGTCAACTGATCCCAG

GATCCATCAGGGATGTGCACCTTGGTGCCTGGTGTTTGCCATAAGGCTTCTCCAGGGCAAATG

TTGGCTGCCCTACAACGGCCATCAACAGGCAGAGTGGTCCCATTAGTATGGCAGGGCGTGAC

ACAGCACAGTCCCCCGTGACTCAAGCCTAGTCCCTGTCTCATACTGGAGGAATGGGGAGCTAA

GGACAGAGCTCCGAGGACATTCCCCCTTAAAGGAATGAGGACACAAGAGAAAGCTCACAGGT

AGTCCATGGGCCAAGTGCAGAGGCAGACAGCCCTAAGCCACGATTGTCTGCGGGGTTTGGCC

CCAGTGAAGTAGTCAGGTAGGGAAGCCTAGGAGCCCCTGGGATGATTGACAGGGCAGAGTTT

GGACCTGGGGTCAAAAGGAAAGAGGAAAAGTGGGTCAGGAAGCACCTGGGTCCCCAGAGCA

GCCCCGAGTGAGTTGGAGCAGGCAGCAGCCGGGGAGGCCACAGTGGAGGCTGCTGGGCCTGG

GATACATGCCACCCCCTGGGAGCAGGACCACAAGGAGGCCTTGCCTCCTCTCACACCTGGTCC

TGCCAAGACCCTGCCTTTGCTTTCTCACTGCATCTCCTTGAAAAAGCAGTGGGACTGTGTCAG

GTTCTGGCTCTACCTCCCAGGCACCACATCTCGGCAGGTAGCCTCAGTGCCGTCCACCTGTGTC

CCTGTTCTCCTTGTCGTTCATACAGGATCATGCATGTGCTGTGCCTAGCACACATTCTTGGCAC

TCACACTGCTGCCTTTTAGCTCTCATCATTTGCCCTCAGAGATCAACCTGAGCTGTGCCCACTG

GGGCGCTCAGAGCAGACCCTGAGCCCCAACACCCAGGCTCCCTGTGCACCTGAGCCTGCCTCT

GCCTGCCACGTGCCCCCAGGCCAGTCCTGGTGGCAGCAAGGATCCGCAAGCTCTCCCCTTTCC

TCATCCTCTGCAAAGCTCTGAATCATCTTTCTCAAAACTTGTTCTGGGAATTTGCTCCGTTGCC

CCAGTTGAGCATGTCAAGCCCGGCGGCCCAAGGCTGGGGTGAAGCAGCGTGGCACGTCACTT
CCCTGGGAACAACTCACACATGGATTGGATTTGGGTCCAACATCCTCTGCCAGGGAAAATAGA
AGCCATAAGAAAACAAAAAAGGAACAGAAGGAGGCTTTTCTTCAGTCACAGCGAGTCACCAA
CAAAAACATGTGCAAAAGCTCTCATGGAGAGCTGGGCCACAAGGAGGGCCATGATGTTGGGG
GCCCTCTGACACCAAGGGTGTGGGCAGGTGGATGGGAGGCAGCTGCCCTCCATGCCAGGCTG
ATGTGCCTCCCTTTGGGTGGTGGGGCTGGGACTCCCACTCCACTTGAAGACCTGCACCAAAA
GTCCTTTAGCCCTGTGCCCAGGCTCTGCCACGGGGCCGGTGAGGGGACTTCTCCCCTCTGCTG
CCAGAGTGAAGCCAGTCAGGGGGATGGGAGGCTTGTAGCCAAGAGCACCTAGTGGCTTTCAG
GGTCCCTTACCCCTGCCACTTAGCAGGGTCTGCACCTGCATCCAAGTGTTCTCCTGGGCTACAG
TGGGGGGCTGGTAGACACTCTGGTGATCCACTTTCAGCTTCCCACATGGATGTGGCAGGGACT
GCTTTGGCATTTCCCTACCCCAAGGGACAGCCACTGCGGCAGGACTGGGCTGGGGAGGGTGG
GGCCTGCGCTGGGGAGGGTGCCCCCTGTCCCTTGCTGCTGCTGGAATGGGAAGGAGAGTTGTT
GAGAGAGCCAGAACTGTCCAAGGGTGGAAGCTGGCGAAACTGACCTGCAGGGAACAGGGAG
ACAGGGAGCATGGCCCAGTGAGTAGGTCCTATGTAGCTCTGAGGCCATCAACCCTGCCATGA
GGGCTGAGACCCCAAGAGAGAAGTTGAGGTTGGGTCAGGGGCCTGTTAGTGCCAGCTGAGGA
GGGGGACAGGCCAGCCTCCTCCCACTGGGACCCAAGCTATAGCTCCTGAGCCTCCAGAGCTGC
CTGGTGCCTCAACCTGGTCAGAGGTGGAAACTCACCTGCCAGCAGGCCCAGTGTGCCTGAGTT
CTGACTGTGGGGATCTGCAGGGCACAGAAGGATAAGAGGTCATCAGGGCCTGGGGACAGGCA
GGAGTGGCAGGGTCTGGGAGGCTGGGAGCAGACCCTCCCAACCTGCCCCATGGCCTCCGTGG
CCCCCAGGACCCCCATGGCAGCAGCTCAGACACGGGTTGTGCCTCAGAAGGAAGTGAAGCTG
TGTGTACCGAGATGGCCCAGCAAACCCTTTGTATGTAAACTTCCGCCACAGCCCAGCTGTCCA
GCACCAGCATGTGTATCTGGGGGAGGGGGATAAATAGAAGGTCTGGGAGGCCTGGGATCTGG
CCAGCAGGCTACTGGGATCACAGATGCCAGCCCCTCCATATCTCCGCTTGAGTCCTGGATCTG
CCTCCTGGGACCAAAGGGGAAAGGACCAGGCTAGGCTCCTTCCTTTTTGTTCTTCCCTCTTGGG
GGAGGCTCCTAGAAACTCCCCCTTCTCTGCCGCCCAAGTGCCTGGATATTACCAGTGGGGTTA
GCCTGTTTGGGCCCACAAGATGGGATGGCTCCCAGAGCCATGGGACCTGAGGTCTCCCAGAC
AGTGTCTAGCCACCCTCACAACTGGCAGAACAATTTCCTTGGTTTTCAACAACTTGAAAAACA
TATGTGATTTTCCACAGTCCGGTGCTTCTCAGGCCTGGCTGCTGAGTGAGCAGAGTTCATGCTG
AATTCCTTCCACTCACCACAGGGCAGACAGCAAGCCCAGCTGTGGGACTCGGTTGGGGTGG
GGGTCACCACAGCAAGGCGCGGGGAGTGGGGAGGGGGGCAGGCTTCCAGCACTGATGAGTA
ATTCTGCTGCCCGAAGATCTGGGAAGAGGGCATGTGACAACTTAGTGCAACAATCTGCCCAGT
GTTAGGTCAGAAGGAAGGAGAGGTCGTTCAAAATGGAGTCTGGTGGAAAAAATAATGTTTGG
CCCCACCTCATACCTCCCTCAAAATTAACTCCAGATTAATGAGGTAGATGTTAGAAGAGGAAC
CAGGGAAGGACTACAAGAAAATATGGAGTCTTTATTTACATTGTGAGGTTTTCTTTAGGTTTT
GTTTGTTTTTGTTTTTGATATGGAGTCTCACTCTGTCACCCAGGCTGGAGTGCAGTGGTGCGAT

CCCGGCTAACTGCAACCTCCGCCTCCCAGGTTCAAGAGATTCTCCTGCCTCAGCCTCCCAAGT
ATCTGGGGATTACAGGCACATGCCACCATGCCCGGCTTTTTTTTTTTTTTTTTTTTTTGTATTT
TTAGTAGAGATGGGGTTTCACCATGTTGACCAGGCAGATCTCAAACTCCTGACCTCAAGTGAT
CCACCCGCCTCAGCCTCCCAAAGTGCTGGGCGCCCGGCATGTGTGCCCAGCCTATATTGACAT

5 TCTTGATGGAGAAGTCTCTTAAGGAAGGACAGAGAAGTTTGGTTGCATAAAAGTTTTTACCTT
CTGTACATCAAAATATACTGAAAATGAAAATAAAGAGCAAACAAAATACTGAGAAAGAATGC
AGTGCTTAGAGAGCGAACATTCCTGGCCTCCTGTAGTTTTAGGAAGCAGCTGTGGCCTCAGAC
CCATCTGCTGTGAACCTCTACTCCATATTTATTGCACTTTCTGTCTGTGAGCGTCGGTTTCTCTC
CTCTATAACAATAGGATAATAATGACACTACCATGCCTTGCAAAAATGCTACAAGGGTTCACT

10 GAGATAAATCTGGAGAGTCATGCCTGAAAAATAGTAAGTCGTTGATAAAGGGAAGCTGCTAT
TAATAAATAAAGCTTTTTCTTTTTTTTTTTTTTGAGATGGAATCTCACTCTGGCGCCTAGGCTGG
AGTGCAGTGATGCAATCTTGGCTCACTGCAACCTCCGCCTCCTGTGTTCAAGCAATCCTCCTAC
TTCAGCATCCTCAGTAGCTGGGACTACAGGTGCGCACCACCATGCCCGGCTAGTTTTTTACATT
TTTAAAGCTATTAATAGGCCAGCCACAGTGGCTCATGCCTATAATCCCAGCACTTTGGGAAGC

15 TGAGGCAGGTGGATC

The adipocyte enhancer binding protein 1 is 16,000 base pairs in length and contains 21 exons (see Table 3 below for location of exons). As will be discussed in further detail below, the human AEBP1 gene is situated in genomic clone AC006454 at nucleotides 137,041-end.

20

POLD2 has an amino acid sequence depicted in SEQ ID NO:4:

MFSEQAAQRAHTLLSPPSANNATFARVPVATYTNSSQPFRLGERSFSRQYAHIYATRLIQMRPFLE
NRAQQHWGSGVGVKKLCELQPEEKCCVVGTLFKAMPLQPSILREVSEEHNLLPQPPRSKYIHPDD
ELVLEDELQRIKLKGTIDVSKLVTGTVLAVFGSVRDDGKFLVEDYCFADLAPQKPAPPLDTDRFVL

25 LVSGLGLGGGGGESLLGTQLLVDVVTGQLGDEGEQCSAAHVSRVILAGNLLSHSTQSRDSINKAK
YLTKKTQAASVEAVKMLDEILLQLSASVPVDVMPGEFDPTNYTLPQQPLHPCMFPLATAYSTLQL
VTNPYQATIDGVRFLGTSGQNVSDIFRYSSMEDHLEILEWTLRVRHISPTAPDTLGCYPFYKTDPFIF
PECPHVYFCGNTPSFGSKIIRGPEDQTVLLVTVPDFSATQTACLVNLRSLACQPISFSGFGAEDDDL
GGLGLGP

30 and a genomic DNA sequence depicted in SEQ ID NO:8..

CCCTCCTCCATCCCTGCCCCAACACCCTGAAGACCCTGGATGCAAACAAAGGCCCGAGGGAG
CCTCTTCCCTCGCAGTGCAGGCCTCACCTGGGGCTCAGAGTCAGAATCTGCATTTTATTCCCTA
GGACAACCTCTAGTCAGGGCAGAGGCCGGCTGTGCTGCCCAAGTGCCCTAACCCTAGCTTTGA
GGCACCAGAAGGGCAAATGCAAATTAAAAATGAGAATAAGTTTATTCTCCTTGGTGAAAAAA

35 AAAAAAAAAGACTTTCCCCTCTCCTTTTTCTTTAGAAAATCTATCATTGCAAGTTCCTTCCTGG

57

ACTTTTTTTATGTAGATCTGTTCAAAAGCTAAATAAGCCTCTTTCAAGTTTCACATCCCAGGAA

TGTCTCCTTAAGGACCTAGGAGCCACCATTTGAAGTGTAATCACCAAGGGAGATACATCCTTA

TCTCCCAGTTTCCGTGGGCAAAGGGGAGCCTAACTTTAGCCCGGTGCCTAGCTCAAGT

TGCAAACACACTTCCAGTCTTAAAGGAATGAATTTATTTTTTTTCCTTTAGGCAAACCCAGGTA

5   GCCACCACAGTTACCTGGGGATTCACAGAGAACTGTGTGTGACCACTGGTGCTGTCAAGTCCT

CTTACCTGAGCACCTGTGACGTTTCCCTTGAGAACGTGTACGGGATGGGTTGCACCTGGTTAT

ATACAAGCGTGAGACTTCTTTCTGCCTTTGTAATTTATTAGCAGATTATCTGTGATGAGC

ATCGCAATCTGTTTAATGCCTATTCAATAATTAAATTTTTCTTTCTCTTCTTTTGTGGAAAGGTT

TTCTGCATTGGCAGGAGATTTTTGTTTTCGATTATGTCCCCAACATGCCTGATGTTCCACCCCT

10   CAAGAGCCTCAGCCTTGCCCAGGGAGGGCATGGGGGTGAGTGGCCTCTCCCACAGAGAGTGC

TGGCCAAGTTGGCCCAGGTGCGCAGCAAGGGCTGCTGCCCAAAGGCTCCCTCCTGGTTG

GCATGGGTCGGGACCCTGTTGTGTTGTGTTTTCGCTCTTTTTCGTAGAGTTCAAGGGGGTCCTG

CTATGTTGTCCAGACTGGTCTTGAACTGACCTCAAGGGATCCTCTCGTCTCAGCCTCCCAAAGT

GCTGGGATTACTGTGCCCAGCTTTGTGTTGTATTTTCTGATCTTATCCTGCAACCTCTTGAGCC

15   CCCAACCTGGGCCCCAGTTCCTGCTGTGCCCCAGCCTGCCAGCCCTCTCTCTCTGCAT

ATTCTTTCTTTAGCTGAGTTAACACCACTGATAAGGTTAAAGACAGGCTCTTAAATTTCTGCCC

TGGCATGAGAAATATGTGACCCACATGCTTCTCCAGCTTAGCTGTCCAGTGTAACTGTCAGGG

ACTGATGGGCGCGTGCTGGCCCACAGCCCACCTCAGTCCTGACCCTCCCTGACAGGCTGAGAG

AGGCCCCAGCCTGAACCTGGACTCCCCCATGTTCTGATATTCCTGCACAAGAGTGCAGAG

20   GCCTGGTTAAGCTGGAGAAACATAAGGAATAGGTAGGTCTGCACACACTCACCTCTTCCTTTG

CAGTGAACCTTCTAGAATCTTCTAGATGGAAAAGCTGGGGGTGTGGAGGTGTAGGGATAGGA

CAGCTGGGGGAGGCCTTGGCCAAGGTCAAGGAGTAGGCCCAGTCTCCCTCTCTGTGTGCCTGT

CTGGGACTCGGTTTCCTGTCTGTGAAGCAGGGCTGGACGGGATATTGACAGCACCTGATGGTC

ATTGAGCTCCTCTGCCCCAGGCACTCAGCTGCTGGGCACAGTGCACACGTGGCAGTCCGGTGC

25   CCTCTCACGCTCCGTGATGACTGAGTCTGTAGTTACACCCTGGCCTCAGAATAAAGACTACA

CTTTCTGCCTCCCTCACTGGCAGGTATGACTAGGTGTGGTGGCAGTTTTCTCCTTAAGAGACAG

ATGTTTGTGCCTCCCTCCAACCCGCTGGCTAACACCTAGCTGGCACACAGCCTCCTGGGGCTA

TGAAGATGAGGGCCACAGCCACAGGGTGGGGGAGCCGTGAGCTGGGTCTGGCTGCGTCTCTG

ACATATGGGGGCATCACACATCACCTCTACCTCCCATCGAATGCTACACGAAGAGAACAAACT

30   CCACCTGATGGAAGCTGCTGTTGTTTGAAGTCTTTCATGCTCACAACAGAACCTAACCCCAAC

CAATACAGTATGAGTATTGGCCCCACGTGGTTAAGCAAGCTGTCCAAGGTTACACACAGCTGG

GAGGTGGTGGAGCTGGGTTTGAGCCTGTTATTGACCTTTGTGCAGACAGACCTCAGAGCAGAG

CACAAGGCAGCAAGGCTGTGGGTCTGGGGCTCCCTCTCCAGGAGAATCAACTGGCTGCACAC

AGCCTGGAGAGCCCATGGGCAACCTGAGTCCTTGCACCTGGAAGTTTCTGTGTCCCACACATA

35   TCCAGGAGCTTAAAATGAAGATGTCTGAATTACCCAACCTCTTGATAGCACCAACCCAACCTT

CCCAGCCTCCTCTTCTGAGGTCAGCCCAGAGCAAGCCCCTTGCAAAGCTGATTTAACTCAGAA

CCACTGGGCATACCCACAGGGCAGTGACCCTGCAGCCCTCGATCAAATGTGCAGATGGACTTG

GGGGTGGGCTGGTACCCCAGATGGCCTCATTCTCCCAGGGTTGCAGAGCCCCTGAAAGCCACA

GCCCTGTGTGCACACCACTGGGGAGTCATCACAGGATACTTCAAGAATTCAGTGCCAGGCAA

5 GGTGGCTCATGGCTGTAATCCCAGCACTTCGGGAGGCTGAAGCGGGCAGATCACCTGAGGTC

AGGAGCTAGAGACCACCCTGGTCAACATAGGGAAACCCCATCTCTACTAAAAATACAAAAAT

TATCTGGGCGTGGTGGCGGGTGCCTGTAATCCCAGCTACTCAGGAGGCTGAGACCGGAAAAT

CGCTTGAGCCTGGGAGGCAGAGGTTGCAGTGAGCTGAGATTGCACTGCTGCACTCCAGCTTGG

GGGACAGAGTAAGACTCCATCTCAGAAAAAGAGTTCTGTGTATCATTTAATGTGGAGATCCT

10 CCCATCACGAGGATGAGGCTGTTTCTCTACTCCCCAGATCTGGGCTGGCCTGTGGTTTGTTGAC

CTCAGCCTTGTAGTTCTCACTTTCCTGGAACCTGAATGCCACCACGCGACATCCATAAGACAA

AGCCCAGGATAAAAGATCACTTGGAGAGACAGGCCTGGCCTGGCACCACCCCGGCTGAGGCT

GGACCCCTGGGAAGGAGACTCTGATGGACCTCCAGACCCAGT

CAAATGACCACTTCCAAGGTCAGGCAAGAAGGGACAAAGAGCCACTGGCTCAGCCCACAGCA

15 TCTGAGAAATAAGAAACCGCTGCATTTTTTGAGCCAGTAAGATTTGACAGGTTTGTTTTGCAG

CAATAGATGAGTGGTACCTCATCTTAGCCCATGTTCTGATGAAGACAAACAGTAGCATTGACA

AAGTTTTAAGAAAAGTTAACCAAAAACTGGGATTCCTTTCTTCATTTTGACCCTTTGTTACAAG

AAACAGAGGCCCACCCCACCAGACTCACTGTTCACTGGTCCCTGAGTGCCTGTGAGTCTCAGT

GGGAGTTACCTTGAGACCAGCCCTTCTGAGTGGAGGGTGCTGGGTGCTGAGGTCAAGTCGAG

20 CTCAGTCCAGGCTAAAAGGAGAGCAGCTCTGGCCAGGCTGTCAGGGCTGTGGCCTCCCCAAG

AACCTCCTACCCTGGCCCCTCCAGGCTTTGCTGCTATGGTTGTGTGAGGGGAGTTGCTGTCCCA

GCATTCTGGCCCCCTTGCCCCCAGCCCCTCCCTGACCTCCACGGGCTTCAGGCCTCAGTCCAGA

GTCACCTCCTCTAGGAAGCCATCCCCCAGTGCAAGTCTGGGCAACATTCCTCCTTGCCTGGCC

CACCTGCTCACTCTCATGCTATGGCTTTCTGTAAGCAAACACAAAGATAGGAACAACTCTGTC

25 CCTGGCACAGAGCAGATGCTCTGGCAATATCTCATGAGTGAATGAAGGCACATGACAAACCT

CCAGACCTGTGGAGACTGAAGGCTGAGAGCCTTTATAGATGCTGTGGGGCCGAGGAGTTTGC

CAACTACAGCAGGTCATGCCCAGAGGTTTCTCTCTGGGTAGCAAGGTGTGTCTCCCACCAAAG

GCCATTGGCATGGGGCCCGCCCTGCTGACCCGAGGCAGTGCACAGCAGAGGCCAGATGCAGT

GAGAAGGAGCCTCTCCTTGGCCTGCTGTCTGCTGCCATGCCTGTGGGGGCGTGGACACAAGTG

30 TGTGGCATAGAAGGTGGTGTGGCAGGTGAGAGGTTGGGGGTGTGTATGTAGCAGGTGTCTGT

GTGTGTATGTGCATGTGGGGGTGTGTGTGCATGCATGTGTGTGTGTGCATATGCACGTGTGTG

CATATGCATGTGTGTGCATGGAGAGAGAAGACCTCCTCTTTCTGGCCCCTCTCCTAGCTGCCCC

CCTCCCTCCTGCTGCCAACACACTGTCAACCCTTCACTGTCTTTTTCCTTGGGACTCGTTGATCT

GTCTCTACCATCCCAGGTGTCTGGAGCAGCCTCTAACCTTCCATCTGCCAAGGTACTTCAGCCC

35 CACCCCTCCCAGCTGTGGAATGTCCCCTAGGATGTGCCACTGACACAAAGAGCCACACAGCTC

CAAAATAGAATATTATCTAACCCACTGCTCCCTTTGCTGTCAGCAACACCTCCACCATGCTTCT

CCCAGGACCCCCCTTGAACTCTCTGCTTCCTCCCTGAGGCCAAAGGAAAGACAGGAAAGGGG

CCACCTTCCTGTCCTTGGGTCCCACAGAGATGTATCCTTGTAATGAAACCTACTTTATGCTTGA

GTTGTATCCAGTTAGTTTCTGTGGCTTGCAATCAAGACCCACACCCACCTCAACCCAGGCTCTA

5   GAGAGTAGACCCTTGTTTTTGCCTGGCTTGGGTCGACCTGGCACCTGCCAGGGTCCCAGCCTC

TGAGTCAGCCCACCTTGCCCTCATCGGTGCCACCTCCAGGCGGCTGT

ACATAGACTCTGGCTTCTGCCCTGGCCTGGCCTCTGGGAACTGCAGCTGTCTGCTTCCATCCTA

TGTGGATGGTGCCTGAAAGTGAATAGGGATCAGTTACCAGCCCAGTATCTGTCCCCTTCTCAA

TAGCACTGATTCCTATGGGGAACTGCTTTTCTTGGACTATGTATGGGTTTGGTGGGAGGGTAG

10 TTCCTGTAACCAACCCTACAGGGTGTAGGAACCTAGACTCTCAGCAACATAACAGGCAGC

AGGCTCCCAAGCTAAGTCTGGCCAGCTGGGCCACCTCTCCCAGATTCTGTTTCATGAGAGCAT

CATCCAAGAGCAGTGGGAACACTGGGGACGGTCCAGCCTAGGACTGGTATGCAGATCAGAGA

ATCCCAGATAGAAGGTGATTGCTGTTCTTCCAGTTTCTTGGCCCTCCAGAGCAACCATACTTCC

CATCTGCCCCAAAACCTGATCCTCCAAACTCCCACCATTTCTGTGCATCCCCAATATCTAA

15 TAGATCAACTGCCTTTCATTTACATTTGTCACAACCAAATGATACACCTGCCCTTCACCCACTA

CTGAACTGCAGCTGGGTTAGTCCAAATTCAGGGCCCACGTGTCATTTCAAGCCTGTCTTGAAT

AATGTACACCTTCCTGCAATGTGAGGATGGCCACCACCTTGGTCTTATACCCACGGGTGTCCT

GAGCTACATTTCTCATAATCAAAAATAAACTCAACACATCACTCCAGCCTGAGCAACAGA

GCAAGACACTAGCTCTAAAAATAAAAAATAAAAACAAACAAATGAAAAACCCAGCAAACTT

20 GGGGAAAGAGGAAGCACCTGATTTCCAGAGTTTCCACATCATGAGATGCAAATGTCCAGTTTT

CAACAACAACAACAACAAAAAAAAAATCACAAGGCATACAAAGAAATAGGAGACTAAG

ACCCACTCAAAGGAAAAGAATAAATAAGCAGAAGCCATACCAGAGGAAAACCAGATGGCTG

ACTTACTAGACAAATACTTTAAAACAACTGTCTTAAAGATGCTTGAAGAGCTAAAGGAAAAT

GTGAACAAAGTCAAGAAAGTGATGGAACAAATGGAAATTCCAATAAAGTGATAGAAACTTT

25 TTGGAGTTTTTTTTCTTGGTAGCAAAAAATTATGAAGCTGAAGAATACAATAAATTCCCTAGA

GGGCTTCAAAGGCAGATGTAAGCAAACTTGGCCAGGTGCAGTGGCTCATGCTCATAATCCAG

CACTTTGGAAGGCTGAGGCAGGAGGATTGCTTGAGCCCAGGAGTTTGAAACCAGCCTGGGCA

ACATAGAAAAACCCTATCTTTAAAAAAACTTATATAAAATTTAAAAATTATAAAATTTATTTA

AAAAATCAGCAATTTGAAGACTGGACAGGGAAATTATCAAATTTGAGGAACAGAAAGGAAA

30 AAGATGGAAGAAAAATAAACAGAGCCTAAGAGACCTGCGGGACACCATCAAGCAGACTAAT

ACCCATTGTGGAAATTCCAGAAGAAAAGAGAGTGAAGGACCAGAGAGATTATTAGGAGAA

ATAATGGCTGAAAATGTCTCAAATTTGATGAATGACATGAATATGAACATTCAAAAATCTCGA

CAAACTCCAAGTAGGAAAAACTCAAAGATACTCATACTGAGATTCATCATAATCAAACTGCTG

AAAGCCAAAGACAAGGAGACAATATCAAAAGCTGCAAGAGAGAAGTGACTCATCACATACA

35 AGGGATCTTCAAAAAGATTATCAGATATCTTGGCTGGGCACGGTGGCTCACACCTGTAATCTT

60

AGCACTTTGGGAGGCCGAGGCAGGTGGATCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGC

CAACATGGCAAAAACCCATCTCCATTAAAAATACAAAGATTGGTGAGGCATGGTGGTGCATG

CCTGTAATCCCAGCTACTCGGGAGGCTGAAGCAGGAGAATCACTTGAACCTGGGAGGCGGAG

GGTGCACCAAGCCAAGATCGTGCCACCACTGCACTCCAGCCTGGGTGACAGAGTGTGACCTTG

TTTCAAAAAAAAAGAAAAGAAAAGAAAAAAAGATCATCAGCTATCTCATCAGAAACCT

CAGAGGCCAAAAGGCAGTAGATTGATATATTCAAAGTGCTAAAAGAAAAAAATAAATCTGTC

AGCTGAGAATCCTGTATCTGTATCTCACTTAACCATTATTTTAAAATAAGGGAAAATGAAGAC

ATTCCCAGATAAACACAAGCTGAGGGAGTTCATTATCACTAGATCTGCCCTGCAAAGAAAGCC

AAAGAAAGCCTTTCAGGATGAAATGAAGGATACTAGACAGTGACTCAAAGCTGAATAAAGA

GGCCAGGCATAGTGGCTCACACCTGTAATCTCAGCACTTTGGGAGGCTGAGATGGGCGGATC

ACCTGAGGAGTTGGAGACCAGCCTGGCTAATATGGTGGAACCCCATCTCTACGAAAAATACA

AAAATTAGCCAGGTGTGGTGGCACATGCCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAAG

AGAATCACCTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCGAGATTGTGCCACCGCACTCCA

GCCTGGGTGACAGAGTGATACCCTGTCTCAAAAAAAAAAGCCGAATAAACGAATAAAGATCT

CATCTATGGCCGTACCACCCTGAATGTGTCCAATCTCAGAAGCTAAGCAGAGTTGGGCCTGGT

TAGTACTTGGAGGGGAGAAATAACGGTCTATGCTAAAGGAAAATTCAGGTGCAATTAAAGTA

AAATTAATTATATAAAAGAGAATACATTAAAAGCTAGTATTATTGTAACTTTGGTTTGTAATT

CCACCAAGTGGAATTTGTTCCTGAAATGCTAGAATGGTTCAACATAAAAATCAATAAATGTAA

TAGACCACATTAACAGAAAAAAAACCCACACGGTCATCTCAATTGATGTCAAAAAAGTATTT

GACAAAATTCAACACTCTTTTGAAAGAAGAAAAAGCTCAACAAACTAAGAATAGGAGGAAAC

TACCTCAAATAATAAAATCCATAGGCCAAATCCCCAAACTCACAGCTAGCAACATATTTAATG

CTAAAGACTGAAAGCTTCCCCTTTAAGATCCGGAATAAGACAAAGATGCCCACTTTCACCACT

TCTACTCAACATAGTATGGGAAGTTCTAGCCAGAGTAATCAGGTAAGAAAAAAGAAATAAAA

AGCATCTGAATTGGAAAGGAAAAGTAAAATTATTTGTTTGCCCAATACATGTACAATGTTTC

AGGTGAAGGCTCAGAACAGTACAACCTTACCAGCAAGAGTCCTGCTGTCTCTGTGTGAATCCC

AGCTATTACTCACTAGCTACATGATCTCTCTTGCCCTCCCTGCCTCAATTTCCTCATGTGTAAA

GTGGGAGAAAAATAATAGTTCATGCTTCAAAGGTTTTTTGTTTGTTTGCTTGCTTTGAGACAGC

GTCTGGCTCTGTCGCTCAGGCTGAAGTGCAGTGGTGCAATCTTAGGTCACTGCAACCTCAGCC

TCCTGGGCTTAAGCGATCCTCCCACCTCGGCCTCCCAAAGTGTTGGGATACAGGCGTGAACCA

CTGTGTCTGACCCAAAGGATTATTTGAGGAGCAGATGAATTAATGTGTCATAACCTCAAAGCA

GTTGCAAAGGCGTTTAATAATTAAAATATCACATTTTAAATTAAAATATAAGGCTGGGCGTGG

TGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGTGGGAGGATCACTTGAGCCCAGG

AGTTCCACACTAGCCTGGGCACCATTGGGAGACCCTGTCTCTACACACACACGCACACACACA

CACACACACAAACTTAAAGTAGCCAGGCGTGGTGCTGCGCGCCTGTTGTCCCAGCTACTCG

GGAGGCTGAGGCGGGAGAATCACTGGAGCCTGGGAGTTCGAGGCTGCAGTGAGCCGAGATCG

CACCACTGCACTCCAGCCTGGGCCACAGAGCAAGACGCTGCCTCAAACAAACAAACAAAAAC
AAAAATTAAAATATTAAGTAATAATTAACGAGTGTTAATATCCACTCGTTGTGGAGACAAGAC
CTGGACTTAGGAAACAGGCCCAGGGAAGTAGCAGAACAGTAGCGCTAGAGGACGCCTGGGA
GAATCAGCGCGCGGCGGGAAGAGCCCGGGAAGCTTAGTGGGGAAGCGTCTCTTGATGGGGTG
5 AGGAATTCTATAAATTAGTGGAGATGGAAAAAAAAAAAAAAAAAGTATTCCCAAAGTGGGAG
ACAGCACTCAGAAAGACGTGGTGGTAAGAACGAGTATGAGTAACGGGGACAACGAGGACAC
TGGAGATTGGGGAGTGTTGGGCTGGAAGCTGGTGTGCAGCTGTGGGCAAGCTAGGGAGGACC
CCGAAACCGCCAATGCGTTTCCCGGACGCAGACGCTGGCAGGACGGGAGGAACCCCGAGACC
CCGCGCCATCCCTTCAGGAAGAGTTACTTCTCCCCGGCCAAGTTAGTGGGCCTTGGGCCTTCTT
10 TCTGTTGGGATCCTCCTCGCGTGTCGCCATCGCTACAAGTGGGCAGCTCTGCGGGGAAAGCTG
GGACGCTGGGGGCTTCACCAAGGAGGCTGGCGGCCGACCACTGGGAGGTCTGGCGGGGTGAC
GACCACTGGGAGGTTTGGGCAGGGCCTGACGGGGTGACGCGGTCAGCCCACTGGAGGCCGAC
ACCCCCCGTCAGCCCAACCCCTGCACGCGCGGCCGCCAACCAAAGACCCGCGGCGCCGGCCT
GCGAGCCCCCGCCCCGCGTTGCCCAGGAAACCGAGGGTGTGGCTCCGCGTTCTCTGGGCGTCC
15 CAGGGACTGGGCGCACAGTGGTCGGCGGGATGAGGCGCCTGGTGACGGACGGGGCGAGGAG
GGCAGCGATTGGTGAGATTAGGCGATGGGCGGGGAAGCCGCGCGGGGATTAGCGAGTTGCGG
CGATGGGCGGGGCAGGCGCGCGGGGATTGGCGGGATGCGGCGCGCCGCGCGTTGAGTGGGGT
CCAGGGAAACGGGGTCAGCTGGGGGTGGCAGTTCCAGGCCGCGAGGCCGGGCTCCTGGGTCG
GTGGGCTGGTGTCTTGGCGGACGTCCCGCAGCTGCCGCGTGGATCCGAGCCGGGGCACCCGCC
20 GTGACTGGGACAGCCCCCAGGGCGCTCTCGGCCCCATCCCGAGTAGCGCGGCCTGGCTGCTGC
CGCCATCAAGCACGTTCGAGCCAAAAGCTCCTAACGAGTCACTCGTTAGACACGTGTGCGGA
GCCTGTGTCCCAGGCCAGTGCTGTCCCGTGGAGATAGATTGCAAGCCGCTAGGGAATTTTTTA
ACTTTCTAGTAGGTGTACGAAAAAAGTAAAACGAAACAAATCAATTGGAGTAAATCCATAAA
TATATTCAAACTATTATTTCAATTGTATGTGAAAAAATTATTGGGATATTCTTTGTACTATTCTT
25 AGAAATCCATTGTGTGTCCAACCCAAACATCACAGTTGGACTCACCACATCTCCTGTACTTCG
TAGCCCTAGGTGGCTAGTGGCATAAGACACAAAAATCTCAGCTCTCCTGGAGCTTATGGTCTA
GTTGGAGCAGGCAGACAATACATTTAAAATATACAGTTTGTTAGAAGGTAAATGTTGTAAACA
ACAATAACAGTTGAAGTACTGGGGAGAGTTGCAGTTGTAAATCAGATGGGCAGGGCACAAGG
TAACATTTGAGTAAAGATGTAAGAACTTGAAGGAGATGGGCAAGTGAGCTCTATAAGTATAC
30 GGGAGAGGGGCAAGCAAGAGTTCAGAGGCCCCTTGCTGTGGGGAGGGATCCAAGGTGGAGG
AGTGGGAACCAGGAGGGGAGAGGACCAGTGGAGCAGATCTCATAGGCAGTTGTAAGGACTTG
GGGCCTTATTCAATGAAATGAGGACACTTTGGAGAGTTTTGAACAGAGCAGTGACTGATTTAT
GTTTTGGTTTTGGTTTAGTTCTATTATTATTTAATAATAGGCTTATTATTTCACAGAAGTTTTAT
TTAATAAGGCAGACCTCTTGTCTGGAAATGAGACAGGTGCCGGAGAGCTGGATGGAGGCAGA
35 TCGGGAATTCCATTTGGGGCAAACTGAACTTGATTGAGACCCTGGTAGTTGTCCAGATGGAAC

AGGACACCTGAGTCTAGGGTTCGGGAAGAACTCCAGATGGGACAAACACTCCTAGCTTTCCTT

TTCTCTTTTTGGATGACCGCTACAGGGTGAGACATCGGTATCCAGGCACGATAAATTTCCAAG

TGGACACAATGTCTGGTGTCAACTACAGCTGTTCTCCTTCTTTTCCCAGTATCCTTTGGGTGCA

GTGAGACACCAGGAGAGCTGCTGCTTTGGGGGATGGACAGGGGCAGCAGGAATGCCTTTGTG

5   TTTTCGCAGTGAACCTCCTTGGCCTGGGCGAAGCTGTGTGGACCAAGCAAGTCAGGAGTGTGG

CCATGTTTTCTGAGCAGGCTGCCCAGAGGGCCCACACTCTACTGTCCCCACCATCAGCCAACA

ATGCCACCTTTGCCCGGGTGCCAGTGGCAACCTACACCAACTCCTCACAACCCTTCCGGCTAG

GAGAGCGCAGCTTTAGCCGGCAGTATGCCCACATTTATGCCACCCGCCTCATCCAAATGAGAC

CCTTCCTGGAGAACCGGGCCCAGCAGCACTGGGGTAAGTGAGAGTTTGGGAAGGTGCTTCCC

10   CCACAGCATCCCTGAACTTAGAAGTGTTCTGCAAGAGAATGGGAACAGTTTATCTAATTGATC

CCACTTCCTGTTACCTTGGGAAAATTAACCTCTTTTTCCCTCAGTTTCTTCTTAAGATAGTAAC

AAGGATTAAATTAAGTAATTTGTGGGTTTGGAGTTAGTTTTAGTTCAGAGGCTGGTTGGAGAT

GAGGACTTAGTTCTGGCGGTGATGGCGATTACTTCACTGGCAGAGGAAAATGGTTTTCCTATC

TTCAGTGCAGATTATTCAGGTATTTGCCTGTGCTGTAGCCAGAGAGCCCCTCAGTGTGGCAAG

15   CCTGGCGCCAGGCACCAGGAGCCAAGACTGGTGAGGATGCACTCTCTGGTCTCGAGGGGACC

CCCTCTGTTCACTCATGTCTGTTTGCCTCTCCTCCTGGCCCCCATATTTGCTGGCCATGAATTTT

CCTGTCCCTTGGGCCCTCTGTCTTTCCTAATAAAGTGGCCTGCCCAACACAACCCTTGTTCTTT

GCCCCCATTTCTTCCCTGGTGATCTCTCCTGCAGTTGGATTACTCTTGGTGGTGAAGCAGGGAC

CCCCATCTCCCCCTTTGAGTTTATTTGAGTTTTAGGTGCTGCTGCATTCCCCCATTCCTACCACT

20   TACATAAGAGTGGCTTTCCAGGTAATTTTCAAATCCATCTCCTATTATATTTTTAAACTGAGGA

TTTAGTAGGTGAGACCAGGTCTTACTCATTTTTACTGTCCTTGGCACCAGGCAAAATGGATCTC

AGCCCTAGTTGCACATTGGAATCCCCTGGGGAGCTTTGAGAAGCCCATCTCATCCCATGCCAA

GCCAAGATCAATTCTCGTTATAGGCAGGCGGAGAACCCTGGGCCTAGAAATCTAGCTAGAAC

CTCAAATTCATTAGGGATATGTATTAGTCCATTTTCACATTGCTATAAAAAACTACCTGAGATA

25   GGGTAATTTATAAAGAAAAGAGGTTTAATTGACTCACAGTTCCTCATGGCTGGGGAGGCCTCA

GGAAACTTAACAATCATGGCAGAAGGTGAAGGGAAAGCAAGGCTCTTTTACATGATAGCAGG

AGAGAGAGAGCAAGGGGAACTGCCAACCATTTTTAAACCATCAGATCGCATGATGGCTTGAT

CTCACTCACCATCACAAGAACAGCATGGGGGAAATCCACCCCCACAATCCAGTCACCTCCCAC

CAGGTCCCTCCGTCAACACCGTGTGGATTATAATTCCAGATGAGATGTGGGTGGGGACACAGA

30   GCCAAATCATATCAGGATGTTTTCTGTTTTGTTTACCTGAGACAAAGTGCTGTTCACCTCTCCT

CTCCCACATAATCAGGGGCTCCCTCCTGCGGCTCCGGTAGCTTTTCCTCACTTTCCTTTCAGCC

CTCGGGACACCTTCCTTGGCTCCTTTCAGAGCTCAGTTACTACTTGGGCCCAATGTCAATGCCA

CCTTCTAGATTCTTTCCGGCAGCACCTCCTCTGGTCGCACATTTCTCTTCCAGTTATTGGAGCT

GTCAAAAAAGCTCCCCAGTGATGGACGATAGCGATTTCACTGTGCTCACAGACTGGTCAGGA

35   AACCAAACAGCTGCCACAGTGAATGTGTTGATAGCAGCGGGGCAGCAGTAGCACTCGCTCAC

63

AGGCCTGGTGGTTGGTGCTGGCCCCCACCCTGAATACCTACATGTGGCTTCTCCATGTGGCCT
GTGCATCCTCACTGAAGCTCAGCCTGTCTCTCCAAATTGGTCTTTCCACTCACCTGTTCCCCAA
ACCTGCCCAGACCTTCCTGCTGTAGGCTTTTCCCTTCACTTGGCACACTCTTTCCCTTGTCTTCC
CATGGCCCCATCTAAGCCCCACTGTCAGCTGAAGTGTTATATTCTTTGAGGGGCCACCTGAAG

5   CCACCTTGCAATGAGGGCCTCCGTTTTCTACCTCAGCTCACCATTTGTTCACAGCACTTGTCAC
TGTGGCGAGTTACTTGTCTATGGCCTGTTGTCGTTCTCCTGCCTAGACCCAGTGGGCTGAGTGG
GGGCAAGTGTTGGCTTTTATGTCCAGTTTTGATCTTGGTGCCAGCACATTGCCTGGGTGGAAG
CATGTCCTACTATCGGTTACAGGGATGTCATTCTGCCCAGTGCTCAGGGGCATACACTTGGAT
CCCAGTTGTGTGCCCTTGGACACATTGCTTAACCTCTCTGTGCATCAGTTGGGTGATAATATCT

10   ACTCCTGGCACATTTTCAGCGTTGGCTGAGTTACATTACAGTGCTTAGGCCACCTGGGGGAGA
GTAAGAGTGGGATACGTGAGGATGTGGAGTCTGTTGCATTTCTGTCTGCTGCTGGCATCCTTCT
TGTCTTGTTTTGAGTTGCTCGCCTCTGTCTGCTCCCTAGGGCGTAGATTTGAGGAATATTCCTG
GTTCTTCCCAGGCAGCAGGGGCTCAGGCTGTGCTGGAGTCAGCTAGGCTAAGGGGCTGGTCTG
GCATCCGCGTTGTCCTGTCACCTCCTTGGTGTTTCTCCAGGCCTGGATCTGTGCTGTGTGGGC

15   ACCTGTATTCCTCCCTCCTGCCCTCACTGATTCTCCATACCTTTCTTCTCGAGAGTGCCAAGCC
CCTCCCATGTGTTCTTGTTCATACCTAGGATCCCGGGAAGGGGCTGGGGAAGACGGTGCCCAG
GTGCCCTGGGTAAACAAAGCCACCTGACTCCACGGGAATGGAATGGGTGGAGGGGATCTGAG
GTCTGCATTTTGAGTATCTCTGGTCTCAGAGGATGAAGCATTTGGTGGGGGTTGGGGGTGGGG
GGTAGGGTGGAAGAATCTAAAGTCTTAAAAGAAAATGGCAGTTATTTGTGGGACAGGGCTGT

20   GTTGAGACTTGGCATGCTTCTTTTTAAGAGTCAGTGTTGTAATTTAGGTATAAGTGAAGCAGT
ACTTTGTATTAGTTTCCTGTAGGCGCTGTAACAAAGCACCACAAACTGGTTGACTTAAAACAA
CAGACATGGCCGGGCACGGTGGCTCACGACTGTAATCCCAGCACTTTGGGAGGCCGAGGCGG
GCAGATCACAAGGTCAAGAGATTGAGACCATCCTGGCTAACACGGTGAAACCCTGTCTCTACT
AAAAATACAAAAAAAAAAAAAATTAGCTGGGCGTGGTGGCACACGCCTGTAGTCCCAGCTACT

25   CGGGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGGAGGCGGAGCTTGCAGTGAGCTGAGA
TCGCGCCACTGCACTCCAGCCTGGATGACAGCGAGACTCCGCCTCAAAACAAAAACAAAAAC
AGAAACAACAATAACAGAAAAACACAGACATTTACTCTCTGGCAGTTCTGGAGGCCAGAAGT
TGAAATCCAGATGTCAGCAGGATTGGCTCCTTCTGAAGGCCCGAGGGGAGGGTCCTTCCTGGC
CTCCTCCCTGGTGTTCCTGGGCTTGTGGCCGCATCACTCCGCTCTGCCCGTCTTCACACTCCCT

30   CTTGTCTGTGTGTCTGTCTCTCTGTTCTCATGAGGACACTTGGCATCCAGGGCCCAACCACACC
CAGAGTCCCTGGTCTCCTGTGGCTGACTCACTTTTTACTGTCACCGTGAAGTCCAGGGGGTCCT
TGTACTTGATGTTCTCTCCTGGCAAGGCCAGGGCCCTGTGATTGGCCTCTCATGGAGTGCTGG
GCAGGGCCTCCATGGCCTCTGTCGGGCGGGGGGGCTACTTCATCTCTGAGTCTGTACCCCTCG
TGTCCCAGGCAGTGGAGTGGGAGTGAAGAAGCTGTGTGAACTGCAGCCTGAGGAGAAGTGCT

35   GTGTGGTGGGCACTCTGTTCAAGGCCATGCCGCTGCAGCCCTCCATCCTGCGGGAGGTCAGCG

64

AGGAGGTGAGGCAGGGTGCTACACAGTGGGGCCGCCAGGCAGACCTGGCCTCCCACTAGAAC
ACCTCCCTGGAGGTGGGGTTGTGGGGAAGCAGGTTCAGAGACAATGGACTCCAGAGGGGTGG
GGGCTGCGGTGCCAGCTCACTAACACCAGAGCTTTGGTGGGCTCTGGCCCCAAGATTATACCT
CCTGTCTCTGCATTCCAGCACAACCTGCTCCCCAGCCTCCTCGGAGTAAATACATACACCCA

5 GATGACGAGCTGGTCTTGGAAGATGAACTGCAGCGTATCAAACTAAAAGGCACCATTGACGT
GTCAAAGCTGGTTACGGGTAGGGAGCCCAATGAGAGGATGTGGGTGATGCAGGTGAAGAGCC
CAGCGGTGGTGTGTTAGGGATGGTGTGAGTGGGGAGCCTGGGGGGGAGTGGGGGGGTGTGGCC
TGGGCACACGTGTGTTCTTGAGGAGGTAGGTGAGGCTCCAGGCGGTCGGAGGCCATCAGATT
GGGTGAGACCTGGCTGGGAGATGGGTCTCCCCACCTCCATCCAAGGGCAGTGACTCCAGGAA

10 GCAGGCATGCATCCTGGAGTCCTAGGTGAGAATTCACCAATGTGGTTGTGGAGAACTGGCTTG
TTTTGCCCGTTGGGGTGACTGGAAGGAGTGGTAGCACCTGGGGCTCCCTGCTCAGGCCTGATG
CCACTGCTCCCCAGGGACTGTCCTGGCTGTGTTTGGCTCCGTGAGAGACGACGGGAAGTTTCT
GGTGGAGGACTATTGCTTTGCTGACCTTGCTCCCCAGAAGCCCGCACCCCCACTTGACACAGA
TAGGTGAGCAGCAGTTCTCGGGAGCTGGAACCAGCTCATGGTCAGTGGAATCTTTGAGTTGCA

15 CCTAGGAGGGGCTGCCTCCCTTCTCGGCACCCTGGAGGACCCCACCTTCTCCCGCAGGTTTGT
GCTACTGGTGTCCGGCCTGGGCCTGGGTGGCGGTGGAGGCGAGAGCCTGCTGGGCACCCAGC
TGCTGGTGGATGTGGTGACGGGGCAGCTTGGGGACGAAGGGGAGCAGTGCAGCGCCGCCCAC
GTCTCCCGGGTTATCCTCGCTGGCAACCTCCTCAGCCACAGCACCCAGAGCAGGGATTCTATC
AATAAGGTATGGAGCCCACCTGGCTGCATTCAGCCCCAGCCCAGGAGCCTGCAAGCCTGTAA

20 GACCCTCCTTCCCCAGGGCGAGTAGGGTACCCTGTGAGGTCTCGCAGGTCGGTGGGAAGCGCC
CTGCAGTGACTCTGGGGCCTCCTGCAATGGGGCTCCTCATGCCCAGGCCCTCGCTGAGGATGG
TGGGAGGCTTGAAGGGAGTGAGGGTCTATGGGACAACAACTGCATCTTCCAGCTGGTGGGGC
TCTACTCCTCTGAGCCTGGGACTCGCCTGGGCCTGATGGCCTTCTGGGCTTCTATTCCAGGC
CAAATACCTCACCAAGAAAACCCAGGCAGCCAGCGTGGAGGCTGTTAAGATGCTGGATGAGA

25 TCCTCCTGCAGCTGAGCGTGAGCGAGCTGGGGGCTGGAGGGGTGATGGGGATTGCAGTCTTC
AAAGCTGCCACTGGGCAACAGAAGGCAGGCAGGAGGGCAGGGGGAGTGGCCGGAGTTGGTG
TAGGGGGCTCCTTCGGGGCCCTGTGAGCTCTCCCTGCCCTGTGCCTTCCAGGCCTCAGTGCCCG
TGGACGTGATGCCAGGCGAGTTTGATCCCACCAATTACACGCTCCCCCAGCAGCCCCTCCACC
CCTGCATGTTCCCGCTGGCCACTGCCTACTCCACGCTCCAGCTGGTCACCAACCCCTACCAGG

30 CCACCATTGATGGAGTCAGGTAGCTGGCACAGCCACACTTCAGTCTGACCCAGCCTTTTGCCT
CAGGAGGCACAAAGAAGGGAGGGGAGGGAGGGCCCAGGAAGGTGGCAGGGCTGCAGAGGC
CCACCTAGCATCTGTTCCTTCTCTCTGGGGCATCCCCACAAGAGCGCCAGATGAGCTCTGGGC
TGACCACTATGGGTGGCACCCAAAGCCAAGAGTCAGCTGAGCTTTGCCTTGCAGATTTTTGGG
GACATCAGGACAGAACGTGAGTGACATTTTCCGATACAGCAGCATGGAGGATCACTTGGAGA

35 TCCTGGAGTGGACCCTGCGGGTCCGTCACATCAGCCCCACAGCCCCGGACACTCTAGGTAACA

65

GGCTCAGCCATACAGGGTGGGAGCAGAGGGCCAGGAGGCCTGGCAGGACCCTGAAGTGCAC

AGGGTCCCCCTGTGGGTTTGCACTTGCCAGCATTGCTGAGAACTGTCTGAGGAGAAGTTCAGA

GGCTTGGCACCTGCTCTGGAAGCTACTCTGGAATCTTAATTCTAAGGCCAATGGCTGCCCACC

CCAACGGGCAGCAACAGCAGGGCCAAGGTCTTGTGACAATGTCTGGAGGTGCCCCTATTGTC

5   ACACTGGGGGTCTCCTACTGGCCTGCAATGGGAGGAGGGGCTGCAGCCCCACATCCTGTGCA

GAGTGCTAGTGCTGAGGCGGAACCCTCCTCAGAGCTGCCCCTTCTCCTCTAGGTTGTTACCCCT

TCTACAAAACTGACCCGTTCATCTTCCCAGAGTGCCCGCATGTCTACTTTTGTGGCAACACCCC

CAGCTTTGGCTCCAAAATCATCCGAGGTAATTTTTGTCTTCTGGGGGCCCAGGCTGATTTGCTG

ATTTGCTCTCACCTGGGGACAAGGTTCACAGAGAAGAAAACCTGCATTGTGGAGTCCCCCTGG

10   CCCTTGTGGGATGGACAGCTGAGGTCTTCTGCACAGCTGCCATTTCACTGTGGGAGCCAAGCT

GCCTCGCCAGCTGGGCAGGGACTGGAACGGCTCCCAGCCTGTGTGCCTCTCAAGGCTAATCTC

TGGTCTCCTATTGTCACTGCCCCACTGTGTGCCAATGGGGACTCCTGTTTATTTCTGGCAGCTT

CTCTTTGAGGCAGGACTTACTTGGAACCTACAGTGGGTCCTATGTGACTTCTTTGCAGGTCCTG

AGGACCAGACAGTGCTGTTGGTGACTGTCCCTGACTTCAGTGCCACGCAGACCGCCTGCCTTG

15   TGAACCTGCGCAGCCTGGCCTGCCAGCCCATCAGCTTCTCGGGCTTCGGGGCAGAGGACGATG

ACCTGGGAGGCCTGGGGCTGGGCCCCTGACTCAAAAAAGTGGTTTTGACCAGAGAGGCCCAG

ATGGAGGCTGTTCATTCCCTGCAGTGTCGGCATTGTAAATAAAGCCTGGCACTTGCTGATGCG

AGCCTTGAGCCCTGGGCACTCTGGCTATGGGACTCCTGCAGGGGTGCCCACAGTGACCATAGC

CCATGCACCCACCAGCCGGTCTCCCT

20

The POLD2 gene is 19,000 base pairs in length and contains ten exons (see Table 4 below for location of

exons). As will be discussed in further detail below, the POLD2 gene is situated in genomic clone

AC006454 at nucleotides 119,001-138,000.

       The polynucleotides of the invention have at least a 95% identity and may have a 96%, 97%, 98%

25   or 99% identity to the polynucleotides depicted in SEQ ID NOS:5, 6, 7 or 8 as well as the polynucleotides

in reverse sense orientation, or the polynucleotide sequences encoding the SNARE YKT6, AEBP1, human

glucokinase or POLD2 polypeptides depicted in SEQ ID NOS:1, 2, 3, or 4 respectively.

       A polynucleotide having 95% "identity" to a reference nucleotide sequence of the present

invention, is identical to the reference sequence except that the polynucleotide sequence may include on

30   average up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding

the polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95%

identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be

deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides

in the reference sequence may be inserted into the reference sequence. The query sequence may be an

35   entire sequence, the ORF (open reading frame), or any fragment specified as described herein.

As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the presence invention can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred

5   to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are: Matrix=Unitary, k-tuple=4,

10   Mismatch Penalty=1, Joining Penalty=30, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB

15   program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment.

20   This percentage is then subtracted from the percent identify, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence are calculated for the purposes of manually adjusting the percent identity score.

25   For example, a 95 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignment of the first 10 bases at 5' end. The 10 unpaired bases represent 5% of the sequence (number of bases at the 5' and 3' ends not matched/total numbers of bases in the query sequence) so 5% is subtracted from the percent identity score calculated by the FASTDB

30   program. If the remaining 95 bases were perfectly matched the final percent identity would be 95%. In another example, a 95 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not

matched/aligned with the query sequence are manually corrected for. No other manual corrections are made for purposes of the present invention.

A polypeptide that has an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence is identical to the query sequence except that the subject polypeptide sequence may include on average, up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the referenced sequence or in one or more contiguous groups within the reference sequence.

A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Com. App. Biosci. (1990) 6:237-245). In a sequence alignment, the query and subject sequence are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject amino acid sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence due to N– or C- terminal deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for N– and C- terminal truncations of the subject sequence when calculating global percent identity. For subject sequences truncated at the N– and C-termini, relative to the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N– and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N– and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N– and C-terminal residues of the subject sequence.

68

The invention also encompasses polynucleotides that hybridize to the polynucleotides depicted in SEQ ID NOS: 5, 6, 7 or 8. A polynucleotide "hybridizes" to another polynucleotide, when a single-stranded form of the polynucleotide can anneal to the other polynucleotide under the appropriate conditions of temperature and solution ionic strength (see Sambrook et al., supra). The conditions of temperature and

5   ionic strength determine the "stringency" of the hybridization. For preliminary screening for homologous nucleic acids, low stringency hybridization conditions, corresponding to a temperature of 42°C, can be used, e.g., 5X SSC, 0.1% SDS, 0.25% milk, and no formamide; or 40% formamide, 5X SSC, 0.5% SDS). Moderate stringency hybridization conditions correspond to a higher temperature of 55°C, e.g., 40% formamide, with 5X or 6X SCC. High stringency hybridization conditions correspond to the highest

10   temperature of 65°C, e.g., 50 % formamide, 5X or 6X SCC. Hybridization requires that the two nucleic acids contain complementary sequences, although depending on the stringency of the hybridization, mismatches between bases are possible. The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids and the degree of complementation, variables well known in the art. The greater the degree of similarity or homology between two nucleotide sequences, the greater the value of Tm

15   for hybrids of nucleic acids having those sequences. The relative stability (corresponding to higher Tm) of nucleic acid hybridizations decreases in the following order: RNA:RNA, DNA:RNA, DNA:DNA.

Polynucleotide and polypeptide variants

The invention is directed to both polynucleotide and polypeptide variants. A "variant" refers to a

20   polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar and in many regions, identical to the polynucleotide or polypeptide of the present invention.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or

25   deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred.

The invention also encompasses allelic variants of said polynucleotides. An allelic variant denotes

30   any of two or more alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in polymorphism within populations. Gene mutations can be silent (no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequences. An allelic variant of a polypeptide is a polypeptide encoded by an allelic variant of a gene.

The amino acid sequences of the variant polypeptides may differ from the amino acid sequences depicted in SEQ ID NOS:1, 2, 3 or 4 by an insertion or deletion of one or more amino acid residues and/or the substitution of one or more amino acid residues by different amino acid residues. Preferably, amino

5   acid changes are of a minor nature, that is conservative amino acid substitutions that do not significantly affect the folding and/or activity of the protein; small deletions, typically of one to about 30 amino acids; small amino- or carboxyl-terminal extensions, such as an amino-terminal methionine residue; a small linker peptide of up to about 20-25 residues; or a small extension that facilitates purification by changing net charge or another function, such as a poly-histidine tract, an antigenic epitope or a binding domain.

10   Examples of conservative substitutions are within the group of basic amino acids (arginine, lysine and histidine), acidic amino acids (glutamic acid and aspartic acid), polar amino acids (glutamine and asparagine), hydrophobic amino acids (leucine, isoleucine and valine), aromatic amino acids (phenylalanine, tryptophan and tyrosine), and small amino acids (glycine, alanine, serine, threonine and methionine). Amino acid substitutions which do not generally alter the specific activity are known in the

15   art and are described, for example, by H. Neurath and R.L. Hill, 1979, *In, The Proteins*, Academic Press, New York. The most commonly occurring exchanges are Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, as well as these in reverse.

20   **Noncoding Regions**

The invention is further directed to polynucleotide fragments containing or hybridizing to noncoding regions of the SNARE YKT6, AEBP1, human glucokinase and POLD2 genes. These include but are not limited to an intron, a 5' non-coding region, a 3' non-coding region and splice junctions (see Tables 1-4), as well as transcription factor binding sites (see Table 5). The polynucleotide fragments may

25   be a short polynucleotide fragment which is between about 8 nucleotides to about 40 nucleotides in length. Such shorter fragments may be useful for diagnostic purposes. Such short polynucleotide fragments are also preferred with respect to polynucleotides containing or hybridizing to polynucleotides containing splice junctions. Alternatively larger fragments, e.g., of about 50, 150, 500, 600 or about 2000 nucleotides in length may be used.

30

70

Table 1: Exon/Intron Regions of Polymerase, DNA directed, 50kD regulat ry subunit (POLD2)
Genomic DNA

	EXONS	LOCATION (nucleotide no.)	
5			(Amino acid no.)
	1.	11546 ------------ 11764	
		1	73
	2.	15534 ------------ 15656	
10		74	114
	3.	15857 ------------ 15979	
		115	155
15	4.	16351 ------------ 16464	
		156	193
	5.	16582 ------------ 16782	
		194	260
20	6.	17089 ------------ 17169	
		261	287
	7.	17327 ------------ 17484	
		288	339
25			
	8.	17704 ------------ 17829	
		340	381
	9.	18199 ------------ 18303	
30		382	416
	10.	18653 ------------ 18811	
		417	469

'tga' at 18812 – 14

35   Poly A at 18885 - 90

71

Table 2: AEBP1 (adip cyte enhancer binding protein 1), vascular smooth muscle-type. Reverse strand coding.

	EXONS	LOCATION (nucleotide no.)	
5			(Amino acid no.)
	21.	1301 ------------ 1966	
		1158	937
10	20.	2209 ------------ 2304	
		936	905
	19.	2426 ------------ 2569	
		904	857
15			
	18.	2651 ------------ 3001	
		856	740
	17.	3238 ------------ 3417	
20		739	680
	16.	3509 ------------ 3706	
		679	614
25	15.	3930 ------------ 4052	
		613	573
	14.	4320 ------------ 4406	
		572	544
30			
	13.	4503 ------------ 4646	
		543	496
	12.	4750 ------------ 4833	
35		495	468

72

11.   5212 ------------ 5352
              467   421

5   10.   5435 ------------ 5545
              420   384

9.   6219 ------------ 6272
              383   366

10

8.   6376 ------------ 6453
              365   340

7.   6584 ------------ 6661
15   339   314

6.   7476 ------------ 7553
              313   288

20   5.   7629 ------------ 7753
              287   247

4.   7860 ------------ 7931
              246   223

25

3.   8050 ------------ 8121
              222   199

2.   8673 ------------ 9014
30   198   85

1.   10642 ---------- 10893
              84   1

35   Stop codon 1298 – 1300
        Poly A-site   1013 - 18

73

Table 3: Glucokinase

	EXONS	LOCATION (nucleotide no.)	
			(Amino acid no.)
5	1.	20485 -------------- 20523	
		1	13
	2.	25133 -------------- 25297	
		14	68
10			
	3.	26173 -------------- 26328	
		69	120
	4.	27524 -------------- 27643	
15		121	160
	5.	28535 -------------- 28630	
		161	192
20	6.	28740 -------------- 28838	
		193	225
	7.	30765 -------------- 30950	
		226	287
25			
	8.	31982 -------------- 32134	
		288	338
	9.	32867 -------------- 33097	
30		339	415
	10.	33314 -------------- 33460	
		416	464

Stop codon   33461-3

74

Table 4: SNARE. Reverse strand coding.

	EXONS	LOCATION (nucleotide no.)	
5			(Amino acid no.)	
	7.	4320 ------------ 4352	
		198	188
	6.	5475 ------------ 5576	
10			187	154
	5.	8401 ------------ 8466	
		153	132
15		4.	9107 ------------ 9211	
		131	97
	3.	10114 ---------- 10215	
		96	63
20				
	2.	11950 ---------- 12033	
		62	35
	1.	15362 ---------- 15463	
25 | | | 34 | 1 |

Stop codon at 4817 – 19

Poly A-site: 4245 – 4250

TABLE 5: TRANSCRIPTION FACTOR BINDING SITES

BINDING SITES	SNARE	GLUCOKINASE	POLD2	AEBP
AP1FJ-Q2	11			11
AP1-C	15	15	7	6
AP1-Q2	9			5
AP1-Q4	7			4
AP4-Q5	36		5	43
AP4-Q6	17			23
ARNT-01	7			5
CEBP-01	7			
CETS1P54-01	6			
CREL-01	7			
DELTAEF1-01	64	12	5	50
FREAC7-01	4			
GATA1-02	19			
GATA1-03	12			6
GATA1-04	25	6		
GATA1-06	8	5		
GATA2-02	10			
GATA3-02	5			
GATA-C	11	6		
GC-01				4
GFII-01	6			
HFH2-01	5			

	Name	Col1	Col2	Col3	Col4
	HFH3-01	10			
	HFH8-01	4			
	IK2-01	49			29
	LMO2COM-01	41	6		27
5	LMO2COM-02	31	5		7
	LYF1-01	10	13	6	
	MAX-01	4			
	MYOD-01	7			
	MYOD-Q6	32	19	7	12
10	MZF1-01	99	40	15	94
	NF1-Q6	5			7
	NFAT-Q6	43	8	7	8
	NFKAPPAB50-01		4		
	NKX25-01	13	14	5	
15	NMYC-01	12			8
	S8-01	30	4		
	SOX5-01	21	20	4	4
	SP1-Q6				8
	SAEBP1-01	4			
20	SRV-02	5			
	STAT-01	6			
	TATA-01	8			
	TCF11-01	47	28	5	19
	USF-01	12	8	6	8
25	USF-C	16	12	12	8
	USF-Q6	6			

In a specific embodiment, such noncoding sequences are expression control sequences. These include but are not limited to DNA regulatory sequences, such as promoters, enhancers, repressors, terminators, and the like, that provide for the regulation of expression of a coding sequence in a host cell. In eukaryotic cells, polyadenylation signals are also control sequences.

5   In a more specific embodiment of the invention, the expression control sequences may be operatively linked to a polynucleotide encoding a heterologous polypeptide. Such expression control sequences may be about 50-200 nucleotides in length and specifically about 50, 100, 200, 500, 600, 1000 or 2000 nucleotides in length. A transcriptional control sequence is "operatively linked" to a polynucleotide encoding a heterologous polypeptide sequence when the expression control sequence

10   controls and regulates the transcription and translation of that polynucleotide sequence. The term "operatively linked" includes having an appropriate start signal (e.g., ATG) in front of the polynucleotide sequence to be expressed and maintaining the correct reading frame to permit expression of the DNA sequence under the control of the expression control sequence and production of the desired product encoded by the polynucleotide sequence. If a gene that one desires to insert into a recombinant DNA

15   molecule does not contain an appropriate start signal, such a start signal can be inserted upstream (5') of and in reading frame with the gene.

Expression of Polypeptides

Isolated Polynucleotide Sequences

The human chromosome 7 genomic clone of accession number AC006454 has been discovered to

20   contain the SNARE YKT6 gene, the human liver glucokinase gene, the AEBP1 gene, and the POLD2 gene by Genscan analysis (Burge et al., 1997, J. Mol. Biol. 268:78-94), BLAST2 and TBLASTN analysis (Altschul et al., 1997, Nucl. Acids Res. 25:3389-3402), in which the sequence of AC006454 was compared to the SNARE YKT6 cDNA sequence, accession number NM_006555 (McNew et al., 1997, J. Biol. Chem. 272:17776-177783), the human liver glucokinase cDNA sequence (Tanizawa et al., 1992, Mol. Endocrinol.

25   6:1070-1081), accession number NM_000162 (major form) and M69051 (minor form), , AEBP1 cDNA sequence, accession number NM_001129 (accession number D86479 for the osteoblast type) (Layne et al., 1998, J. Biol. Chem. 273:15654-15660) and the POLD2 cDNA sequence, accession number NM_006230 (Zhang et al., 1995, Genomics 29:179-186).

The cloning of the nucleic acid sequences of the present invention from such genomic DNA can be

30   effected, e.g., by using the well known polymerase chain reaction (PCR) or antibody screening of expression libraries to detect cloned DNA fragments with shared structural features. See, e.g., Innis et al., 1990, PCR: A Guide to Methods and Application, Academic Press, New York. Other nucleic acid amplification procedures such as ligase chain reaction (LCR), ligated activated transcription (LAT) and nucleic acid sequence-based amplification (NASBA) or long chain PCR may be used. In a specific

embodiment, 5' or 3' non-coding portions of each gene may be identified by methods including but are not limited to, filter probing, clone enrichment using specific probes and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., 1993, Nucl. Acids Res. 21:1683-1684).

Once the DNA fragments are generated, identification of the specific DNA fragment containing the desired SNARE YKT6 gene, the human liver glucokinase gene, the AEBP1 gene, or POLD2 gene may be accomplished in a number of ways. For example, if an amount of a portion of a SNARE YKT6 gene, the human liver glucokinase gene, the AEBP1 gene, or POLD2 gene or its specific RNA, or a fragment thereof, is available and can be purified and labeled, the generated DNA fragments may be screened by nucleic acid hybridization to the labeled probe (Benton and Davis, 1977, Science 196:180; Grunstein and Hogness, 1975, Proc. Natl. Acad. Sci. U.S.A. 72:3961). The present invention provides such nucleic acid probes, which can be conveniently prepared from the specific sequences disclosed herein, e.g., a hybridizable probe having a nucleotide sequence corresponding to at least a 10, and preferably a 15, nucleotide fragment of the sequences depicted in SEQ ID NOS:5, 6, 7 or 8. Preferably, a fragment is selected that is highly unique to the encoded polypeptides. Those DNA fragments with substantial homology to the probe will hybridize. As noted above, the greater the degree of homology, the more stringent hybridization conditions can be used. In one embodiment, low stringency hybridization conditions are used to identify a homologous SNARE YKT6, the human liver glucokinase, the AEBP1, or POLD2 polynucleotide. However, in a preferred aspect, and as demonstrated experimentally herein, a nucleic acid encoding a polypeptide of the invention will hybridize to a nucleic acid derived from the polynucleotide sequence depicted in SEQ ID NOS:5, 6, 7 or 8 or a hybridizable fragment thereof, under moderately stringent conditions; more preferably, it will hybridize under high stringency conditions.

Alternatively, the presence of the gene may be detected by assays based on the physical, chemical, or immunological properties of its expressed product. For example, cDNA clones, or DNA clones which hybrid-select the proper mRNAs, can be selected which produce a protein that, e.g., has similar or identical electrophoretic migration, isoelectric focusing behavior, proteolytic digestion maps, or antigenic properties as known for the SNARE YKT6, the human liver glucokinase, the AEBP1, or POLD2 polynucleotide.

A gene encoding SNARE YKT6, the human liver glucokinase, the AEBP1, or POLD2 polypeptide can also be identified by mRNA selection, i.e., by nucleic acid hybridization followed by *in vitro* translation. In this procedure, fragments are used to isolate complementary mRNAs by hybridization. Immunoprecipitation analysis or functional assays of the in vitro translation products of the products of the isolated mRNAs identifies the mRNA and, therefore, the complementary DNA fragments, that contain the desired sequences.

79

Nucleic Acid Constructs

The present invention also relates to nucleic acid constructs comprising a polynucleotide sequence containing the exon/intron segments of the SNARE YKT6 gene (nucleotides 4320-15463 of SEQ ID NO:5), human liver glucokinase gene (nucleotides 20485-33460 of SEQ ID NO:6), AEBP1 gene

5   (nucleotides 1301-13893 of SEQ ID NO:7) or POLD2 gene (nucleotides 11546-18811 of SEQ ID NO:8) operably linked to one or more control sequences which direct the expression of the coding sequence in a suitable host cell under conditions compatible with the control sequences. Expression will be understood to include any step involved in the production of the polypeptide including, but not limited to, transcription, post-transcriptional modification, translation, post-translational modification, and secretion.

10   The invention is further directed to a nucleic acid construct comprising expression control sequences derived from SEQ ID NOS: 5, 6, 7 or 8 and a heterologous polynucleotide sequence.

"Nucleic acid construct" is defined herein as a nucleic acid molecule, either single- or double-stranded, which is isolated from a naturally occurring gene or which has been modified to contain segments of nucleic acid which are combined and juxtaposed in a manner which would not otherwise exist in nature.

15   The term nucleic acid construct is synonymous with the term expression cassette when the nucleic acid construct contains all the control sequences required for expression of a coding sequence of the present invention. The term "coding sequence" is defined herein as a portion of a nucleic acid sequence which directly specifies the amino acid sequence of its protein product. The boundaries of the coding sequence are generally determined by a ribosome binding site (prokaryotes) or by the ATG start codon (eukaryotes)

20   located just upstream of the open reading frame at the 5' end of the mRNA and a transcription terminator sequence located just downstream of the open reading frame at the 3' end of the mRNA. A coding sequence can include, but is not limited to, DNA, cDNA, and recombinant nucleic acid sequences.

The isolated polynucleotide of the present invention may be manipulated in a variety of ways to provide for expression of the polypeptide. Manipulation of the nucleic acid sequence prior to its insertion

25   into a vector may be desirable or necessary depending on the expression vector. The techniques for modifying nucleic acid sequences utilizing recombinant DNA methods are well known in the art.

The control sequence may be an appropriate promoter sequence, a nucleic acid sequence which is recognized by a host cell for expression of the nucleic acid sequence. The promoter sequence contains transcriptional control sequences which regulate the expression of the polynucleotide. The promoter may

30   be any nucleic acid sequence which shows transcriptional activity in the host cell of choice including mutant, truncated, and hybrid promoters, and may be obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell.

Examples of suitable promoters for directing the transcription of the nucleic acid constructs of the present invention, especially in a bacterial host cell, are the promoters obtained from the *E. coli lac* operon, the prokaryotic beta-lactamase gene (Villa-Komaroff *et al.*, 1978, *Proc. Natl. Acad. Sci. USA* 75: 3727-3731), as well as the *tac* promoter (DeBoer *et al.*, 1983, *Proc. Natl Acad. of Sciences USA* 80: 21-25).

5   Further promoters are described in "Useful proteins from recombinant bacteria" in *Scientific American*, 1980, 242: 74-94; and in Sambrook *et al.*, 1989, *supra*.

Examples of suitable promoters for directing the transcription of the nucleic acid constructs of the present invention in a filamentous fungal host cell are promoters obtained from the genes encoding *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-

10   amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (*glaA*), *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase, *Fusarium oxysporum* trypsin-like protease (WO 96/00787), NA2-tpi (a hybrid of the promoters from the genes encoding *Aspergillus niger* neutral alpha-amylase and *Aspergillus oryzae* triose phosphate isomerase), and mutant, truncated, and hybrid

15   promoters thereof.

In a yeast host, useful promoters are obtained from the *Saccharomyces cerevisiae* enolase (ENO-1) gene, the *Saccharomyces cerevisiae* galactokinase gene (GAL1), the *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase genes (ADH2/GAP), and the *Saccharomyces cerevisiae* 3-phosphoglycerate kinase gene. Other useful promoters for yeast host cells are described by

20   Romanos *et al.*, 1992, *Yeast* 8: 423-488.

Eukaryotic promoters may be obtained from the genomes of viruses such as polyoma virus, fowlpox virus, adenovirus, bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-B virus and SV40. Alternatively, heterologous mammalian promoters, such as the actin promoter or immunoglobulin promoter may be used.

25   The constructs of the invention may also include enhancers. Enhancers are cis-acting elements of DNA, usually from about 10 to about 300 bp that act on a promoter to increase its transcription. Enhancers from globin, elastase, albumin, alpha-fetoprotein, and insulin enhancers may be used. However, an enhancer from a virus may be used; examples include SV40 on the late side of the replication origin, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin

30   and adenovirus enhancers.

The control sequence may also be a suitable transcription terminator sequence, a sequence recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3'

81

terminus of the nucleic acid sequence encoding the polypeptide. Any terminator which is functional in the host cell of choice may be used in the present invention.

The control sequence may also be a suitable leader sequence, a nontranslated region of an mRNA which is important for translation by the host cell. The leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the polypeptide. Any leader sequence that is functional in the host cell of choice may be used in the present invention.

The control sequence may also be a polyadenylation sequence, a sequence which is operably linked to the 3' terminus of the nucleic acid sequence and which, when transcribed, is recognized by the host cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence which is functional in the host cell of choice may be used in the present invention.

The control sequence may also be a signal peptide coding region, which codes for an amino acid sequence linked to the amino terminus of the polypeptide which can direct the encoded polypeptide into the cell's secretory pathway. The 5' end of the coding sequence of the nucleic acid sequence may inherently contain a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region which encodes the secreted polypeptide. Alternatively, the 5' end of the coding sequence may contain a signal peptide coding region which is foreign to the coding sequence. The foreign signal peptide coding region may be required where the coding sequence does not normally contain a signal peptide coding region. Alternatively, the foreign signal peptide coding region may simply replace the natural signal peptide coding region in order to obtain enhanced secretion of the polypeptide. However, any signal peptide coding region which directs the expressed polypeptide into the secretory pathway of a host cell of choice may be used in the present invention.

The control sequence may also be a propeptide coding region, which codes for an amino acid sequence positioned at the amino terminus of a polypeptide. The resultant polypeptide is known as a proenzyme or propolypeptide (or a zymogen in some cases). A propolypeptide is generally inactive and can be converted to a mature active polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide. The propeptide coding region may be obtained from the *Bacillus subtilis* alkaline protease gene (*aprE*), the *Bacillus subtilis* neutral protease gene (*nprT*), the *Saccharomyces cerevisiae* alpha-factor gene, the *Rhizomucor miehei* aspartic proteinase gene, or the *Myceliophthora thermophila* laccase gene (WO 95/33836).

Where both signal peptide and propeptide regions are present at the amino terminus of a polypeptide, the propeptide region is positioned next to the amino terminus of a polypeptide and the signal peptide region is positioned next to the amino terminus of the propeptide region.

82

It may also be desirable to add regulatory sequences which allow the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those which cause the expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. Regulatory systems in prokaryotic systems would

5   include the *lac*, *tac*, and *trp* operator systems. In yeast, the ADH2 system or GAL1 system may be used. In filamentous fungi, the TAKA alpha-amylase promoter, *Aspergillus niger* glucoamylase promoter, and the *Aspergillus oryzae* glucoamylase promoter may be used as regulatory sequences. Other examples of regulatory sequences are those which allow for gene amplification. In eukaryotic systems, these include the dihydrofolate reductase gene which is amplified in the presence of methotrexate, and the

10   metallothionein genes which are amplified with heavy metals. In these cases, the nucleic acid sequence encoding the polypeptide would be operably linked with the regulatory sequence.

Expression Vectors

The present invention also relates to recombinant expression vectors comprising a nucleic acid sequence of the present invention, a promoter, and transcriptional and translational stop signals. The

15   various nucleic acid and control sequences described above may be joined together to produce a recombinant expression vector which may include one or more convenient restriction sites to allow for insertion or substitution of the nucleic acid sequence encoding the polypeptide at such sites. Alternatively, the polynucleotide of the present invention may be expressed by inserting the nucleic acid sequence or a nucleic acid construct comprising the sequence into an appropriate vector for expression. In creating the

20   expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression.

The recombinant expression vector may be any vector (*e.g.*, a plasmid or virus) which can be conveniently subjected to recombinant DNA procedures and can bring about the expression of the nucleic acid sequence. The choice of the vector will typically depend on the compatibility of the vector with the

25   host cell into which the vector is to be introduced. The vectors may be linear or closed circular plasmids.

The vector may be an autonomously replicating vector, *i.e.*, a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, *e.g.*, a plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively, the vector may be one which, when

30   introduced into the host cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. Furthermore, a single vector or plasmid or two or more vectors or plasmids which together contain the total DNA to be introduced into the genome of the host cell, or a transposon may be used.

83

The vectors of the present invention preferably contain one or more selectable markers which permit easy selection of transformed cells. A selectable marker is a gene the product of which provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like. Examples of bacterial selectable markers are the *dal* genes from *Bacillus subtilis* or *Bacillus licheniformis*, or markers

5   which confer antibiotic resistance such as ampicillin, kanamycin, chloramphenicol or tetracycline resistance. Suitable markers for yeast host cells are ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3. An example of suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take of the nucleic acids of the present invention, such as DHFR or thymidine kinase. An appropriate host cell when wild-type DHFR is employed is the CHO cell line deficient in DHFR

10   activity, prepared and propagated as described by Urlaub et al., Proc. Natl. Acad. Sci. USA, 77:4216 (1980).

The vectors of the present invention preferably contain an element(s) that permits stable integration of the vector into the host cell genome or autonomous replication of the vector in the cell independent of the genome of the cell.

15   For integration into the host cell genome, the vector may rely on the polynucleotide sequence encoding the polypeptide or any other element of the vector for stable integration of the vector into the genome by homologous or nonhomologous recombination. Alternatively, the vector may contain additional nucleic acid sequences for directing integration by homologous recombination into the genome of the host cell. The additional polynucleotide sequences enable the vector to be integrated into the host

20   cell genome at a precise location(s) in the chromosome(s). To increase the likelihood of integration at a precise location, the integrational elements should preferably contain a sufficient number of nucleic acids, such as 100 to 1,500 base pairs, preferably 400 to 1,500 base pairs, and most preferably 800 to 1,500 base pairs, which are highly homologous with the corresponding target sequence to enhance the probability of homologous recombination. The integrational elements may be any sequence that is homologous with the

25   target sequence in the genome of the host cell. Furthermore, the integrational elements may be non-encoding or encoding nucleic acid sequences. On the other hand, the vector may be integrated into the genome of the host cell by non-homologous recombination.

For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the host cell in question. Examples of bacterial origins of replication

30   are the origins of replication of plasmids pBR322, pUC19, pACYC177, and pACYC184 permitting replication in *E. coli*, and pUB110, pE194, pTA1060, and pAMß1 permitting replication in *Bacillus*. Examples of origins of replication for use in a yeast host cell are the 2 micron origin of replication, ARS1, ARS4, the combination of ARS1 and CEN3, and the combination of ARS4 and CEN6. The origin of

replication may be one having a mutation which makes its functioning temperature-sensitive in the host cell (see, *e.g.*, Ehrlich, 1978, *Proceedings of the National Academy of Sciences USA* 75: 1433).

More than one copy of a polynucleotide sequence of the present invention may be inserted into the host cell to increase production of the gene product. An increase in the copy number of the polynucleotide

5   sequence can be obtained by integrating at least one additional copy of the sequence into the host cell genome or by including an amplifiable selectable marker gene with the nucleic acid sequence where cells containing amplified copies of the selectable marker gene, and thereby additional copies of the nucleic acid sequence, can be selected for by cultivating the cells in the presence of the appropriate selectable agent.

The procedures used to ligate the elements described above to construct the recombinant

10   expression vectors of the present invention are well known to one skilled in the art (see, *e.g.*, Sambrook *et al.*, 1989, *supra*).

Host Cells

The present invention also relates to recombinant host cells, comprising a nucleic acid sequence of the invention, which are advantageously used in the recombinant production of the polypeptides. A vector

15   comprising a nucleic acid sequence of the present invention is introduced into a host cell so that the vector is maintained as a chromosomal integrant or as a self-replicating extra-chromosomal vector as described earlier. The term "host cell" encompasses any progeny of a parent cell that is not identical to the parent cell due to mutations that occur during replication. The choice of a host cell will to a large extent depend upon the gene encoding the polypeptide and its source.

20   The host cell may be a unicellular microorganism, *e.g.*, a prokaryote, or a non-unicellular microorganism, *e.g.*, a eukaryote. Useful unicellular cells are bacterial cells such as gram positive bacteria including, but not limited to, a *Bacillus* cell, or a *Streptomyces* cell, *e.g.*, *Streptomyces lividans* or *Streptomyces murinus*, or gram negative bacteria such as *E. coli* and *Pseudomonas* sp.

The introduction of a vector into a bacterial host cell may, for instance, be effected by protoplast

25   transformation (see, *e.g.*, Chang and Cohen, 1979, *Molecular General Genetics* 168: 111-115), using competent cells (see, *e.g.*, Young and Spizizin, 1961, *Journal of Bacteriology* 81: 823-829, or Dubnau and Davidoff-Abelson, 1971, *Journal of Molecular Biology* 56: 209-221), electroporation (see, *e.g.*, Shigekawa and Dower, 1988, *Biotechniques* 6: 742-751), or conjugation (see, *e.g.*, Koehler and Thorne, 1987, *Journal of Bacteriology* 169: 5771-5278).

30   The host cell may be a eukaryote, such as a mammalian cell (e.g., human cell), an insect cell, a plant cell or a fungal cell. Mammalian host cells that could be used include but are not limited to human Hela, embryonic kidney cells (293), lung cells, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese Hamster ovary (CHO) cells. These cells

85

may be transfected with a vector containing a transcriptional regulatory sequence, a protein coding sequence and transcriptional termination sequences. Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

The host cell may be a fungal cell. "Fungi" as used herein includes the phyla Ascomycota, Basidiomycota, Chytridiomycota, and Zygomycota (as defined by Hawksworth *et al., In, Ainsworth and Bisby's Dictionary of The Fungi*, 8th edition, 1995, CAB International, University Press, Cambridge, UK) as well as the Oomycota (as cited in Hawksworth *et al.*, 1995, *supra*, page 171) and all mitosporic fungi (Hawksworth *et al.*, 1995, *supra*). The fungal host cell may also be a yeast cell. "Yeast" as used herein includes ascosporogenous yeast (Endomycetales), basidiosporogenous yeast, and yeast belonging to the Fungi Imperfecti (Blastomycetes). Since the classification of yeast may change in the future, for the purposes of this invention, yeast shall be defined as described in *Biology and Activities of Yeast* (Skinner, F.A., Passmore, S.M., and Davenport, R.R., eds, *Soc. App. Bacteriol. Symposium Series* No. 9, 1980). The fungal host cell may also be a filamentous fungal cell. "Filamentous fungi" include all filamentous forms of the subdivision Eumycota and Oomycota (as defined by Hawksworth *et al.*, 1995, *supra*). The filamentous fungi are characterized by a mycelial wall composed of chitin, cellulose, glucan, chitosan, mannan, and other complex polysaccharides. Vegetative growth is by hyphal elongation and carbon catabolism is obligately aerobic. In contrast, vegetative growth by yeasts such as *Saccharomyces cerevisiae* is by budding of a unicellular thallus and carbon catabolism may be fermentative.

Fungal cells may be transformed by a process involving protoplast formation, transformation of the protoplasts, and regeneration of the cell wall in a manner known *per se*. Suitable procedures for transformation of *Aspergillus* host cells are described in EP 238 023 and Yelton *et al.*, 1984, *Proceedings of the National Academy of Sciences USA* 81: 1470-1474. Suitable methods for transforming *Fusarium* species are described by Malardier *et al.*, 1989, *Gene* 78: 147-156 and WO 96/00787. Yeast may be transformed using the procedures described by Becker and Guarente, *In* Abelson, J.N. and Simon, M.I., editors, *Guide to Yeast Genetics and Molecular Biology*, Methods in Enzymology, Volume 194, pp 182-187, Academic Press, Inc., New York; Ito *et al.*, 1983, *Journal of Bacteriology* 153: 163; and Hinnen *et al.*, 1978, *Proc. e Natl Acad. f Sci.s USA* 75: 1920.

Methods of Production

The present invention also relates to methods for producing a polypeptide of the present invention comprising (a) cultivating a host cell under conditions conducive for production of the polypeptide; and (b) recovering the polypeptide.

86

In the production methods of the present invention, the cells are cultivated in a nutrient medium suitable for production of the polypeptide using methods known in the art. For example, the cell may be cultivated by shake flask cultivation, small-scale or large-scale fermentation (including continuous, batch, fed-batch, or solid state fermentations) in laboratory or industrial fermentors performed in a suitable

5   medium and under conditions allowing the polypeptide to be expressed and/or isolated. The cultivation takes place in a suitable nutrient medium comprising carbon and nitrogen sources and inorganic salts, using procedures known in the art. Suitable media are available from commercial suppliers or may be prepared according to published compositions (*e.g.*, in catalogues of the American Type Culture Collection). If the polypeptide is secreted into the nutrient medium, the polypeptide can be recovered directly from the

10   medium. If the polypeptide is not secreted, it can be recovered from cell lysates.

The polypeptides may be detected using methods known in the art that are specific for the polypeptides. These detection methods may include use of specific antibodies, formation of an enzyme product, or disappearance of an enzyme substrate. In a specific embodiment, an enzyme assay may be used to determine the activity of the polypeptide. For example, AEBP1 activity can be determined by measuring

15   carboxypeptidase activity as described by Muise and Ro, 1999, Biochem. J. 343:341-345. Here, the conversion of hippuryl-L-arginine, hippuryl-L-lysine or hippuryl-L-phenylalanine to hippuric acid may be monitored spectrophotometrically. POLD2 activity may be detected by assaying for DNA polymerase _ activity (see, for example, Ng et al., 1991, J. Biol. Chem. 266:11699-11704).

The resulting polypeptide may be recovered by methods known in the art. For example, the

20   polypeptide may be recovered from the nutrient medium by conventional procedures including, but not limited to, centrifugation, filtration, extraction, spray-drying, evaporation, or precipitation.

The polypeptides of the present invention may be purified by a variety of procedures known in the art including, but not limited to, chromatography (*e.g.*, ion exchange, affinity, hydrophobic, chromatofocusing, and size exclusion), electrophoretic procedures (*e.g.*, preparative isoelectric focusing,

25   differential solubility (*e.g.*, ammonium sulfate precipitation), SDS-PAGE, or extraction (see, *e.g.*, *Protein Purification*, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989).

Antibodies

According to the invention, the SNARE YKT6, human glucokinase, AEBP1 or POLD2 polypeptides produced according to the method of the present invention may be used as an immunogen to

30   generate any of these polypeptides. Such antibodies include but are not limited to polyclonal, monoclonal, chimeric, single chain, Fab fragments, and an Fab expression library.

Various procedures known in the art may be used for the production of antibodies. For the production of antibody, various host animals can be immunized by injection with the polypeptide thereof,

including but not limited to rabbits, mice, rats, sheep, goats, etc. In one embodiment, the polypeptide or fragment thereof can optionally be conjugated to an immunogenic carrier, e.g., bovine serum albumin (BSA) or keyhole limpet hemocyanin (KLH). Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete

5   and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanins, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and Corynebacterium parvum.

        For preparation of monoclonal antibodies directed toward the SNARE YKT6, human glucokinase, AEBP1 or POLD2 polypeptide, any technique that provides for the production of antibody molecules by

10   continuous cell lines in culture may be used. These include but are not limited to the hybridoma technique originally developed by Kohler and Milstein (1975, Nature 256:495-497), as well as the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96). In an additional embodiment of the

15   invention, monoclonal antibodies can be produced in germ-free animals utilizing recent technology (PCT/US90/02545). According to the invention, human antibodies may be used and can be obtained by using human hybridomas (Cote et al., 1983, Proc. Natl. Acad. Sci. U.S.A. 80:2026-2030) or by transforming human B cells with EBV virus in vitro (Cole et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, pp. 77-96). In fact, according to the invention, techniques developed for the

20   production of "chimeric antibodies" (Morrison et al., 1984, J. Bacteriol. 159-870; Neuberger et al., 1984, Nature 312:604-608; Takeda et al., 1985, Nature 314:452-454) by splicing the genes from a mouse antibody molecule specific for the SNARE YKT6, human glucokinase, AEBP1 or POLD2 polypeptide together with genes from a human antibody molecule of appropriate biological activity can be used; such antibodies are within the scope of this invention.

25   According to the invention, techniques described for the production of single chain antibodies (U.S. Pat. No. 4,946,778) can be adapted to produce polypeptide-specific single chain antibodies. An additional embodiment of the invention utilizes the techniques described for the construction of Fab expression libraries (Huse et al., 1989, Science 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity for the SNARE YKT6, AEBP1, human glucokinase

30   or POLD2 polypeptides.

        Antibody fragments which contain the idiotype of the antibody molecule can be generated by known techniques. For example, such fragments include but are not limited to: the F(ab')2 fragment which can be produced by pepsin digestion of the antibody molecule; the Fab' fragments which can be generated

by reducing the disulfide bridges of the F(ab')2, fragment,and the Fab fragments which can be generated by treating the antibody molecule with papain and a reducing agent.

In the production of antibodies, screening for the desired antibody can be accomplished by techniques known in the art, e.g., radioimmunoassay, ELISA (enzyme-linked immunosorbent assay), "sandwich" immunoassays, immunoradiometric assays, gel diffusion precipitin reactions, immunodiffusion assays, in situ immunoassays (using colloidal gold, enzyme or radioisotope labels, for example), western blots, precipitation reactions, agglutination assays (e.g., gel agglutination assays, hemagglutination assays), complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, etc. In one embodiment, antibody binding is detected by detecting a label on the primary antibody. In another embodiment, the primary antibody is detected by detecting binding of a secondary antibody or reagent to the primary antibody. In a further embodiment, the secondary antibody is labeled. Many means are known in the art for detecting binding in an immunoassay and are within the scope of the present invention. For example, to select antibodies which recognize a specific epitope of a particular polypeptide, one may assay generated hybridomas for a product which binds to a particular polypeptide fragment containing such epitope. For selection of an antibody specific to a particular polypeptide from a particular species of animal, one can select on the basis of positive binding with the polypeptide expressed by or isolated from cells of that species of animal.

Immortal, antibody-producing cell lines can also be created by techniques other than fusion, such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein-Barr virus. See, e.g., M. Schreier et al., "Hybridoma Techniques" (1980); Hammerling et al., "Monoclonal Antibodies And T-cell Hybridomas" (1981); Kennett et al.,"Monoclonal Antibodies" (1980); see also U.S. Pat. Nos. 4,341,761; 4,399,121; 4,427,783; 4,444,887; 4,451,570; 4,466,917; 4,472,500; 4,491,632; 4,493,890.

Uses of Polynucleotides

Diagnostics

Polynucleotides containing noncoding regions of SEQ ID NOS:5, 6, 7 or 8 may be used as probes for detecting mutations from samples from a patient. Genomic DNA may be isolated from the patient. A mutation(s) may be detected by Southern blot analysis, specifically by hybridizing restriction digested genomic DNA to various probes and subjecting to agarose electrophoresis.

Polynucleotides containing noncoding regions may be used as PCR primers and may be used to amplify the genomic DNA isolated from the patients. Additionally, primers may be obtained by routine or long range PCR, that can yield products containing more than one exon and intervening intron. The sequence of the amplified genomic DNA from the patient may be determined using methods known in the

art. Such probes may be between 10-100 nucleotides in length and may preferably be between 20-50 nucleotides in length.

Thus the invention is thus directed to kits comprising these polynucleotide probes. In a specific embodiment, these probes are labeled with a detectable substance.

5   Antisense Oligonucleotides and Mimetics

The invention is further directed to antisense oligonucleotides and mimetics to these polynucleotide sequences. Antisense technology can be used to control gene expression through triple-helix formation or antisense DNA or RNA, both of which methods are based on binding of a polynucleotide to DNA or RNA. A DNA oligonucleotide is designed to be complementary to a region of the gene involved in transcription
10   or RNA processing (triple helix (see Lee et al., Nucl. Acids Res., 6:3073 (1979); Cooney et al, Science, 241:456 (1988); and Dervan et al., Science, 251: 1360 (1991)), thereby preventing transcription and the production of said polypeptides.

The antisense oligonucleotides or mimetics of the present invention may be used to decrease levels of a polypeptide. For example, SNARE YKT6 has been found to be essential for vesicle-associated
15   endoplasmic reticulum-Golgi transport and cell growth. Therefore, the SNARE YKT6 antisense oligonucleotides of the present invention could be used to inhibit cell growth and in particular, to treat or prevent tumor growth. POLD2 is necessary for DNA replication. POLD2 antisense sequences could also be used to inhibit cell growth. Glucokinase and AEBP1 antisense sequences may be used to treat hyperglycemia.

20   The antisense oligonucleotides of the present invention may be formulated into pharmaceutical compositions. These compositions may be administered in a number of ways depending upon whether local or systemic treatment is desired and upon the area to be treated. Administration may be topical (including ophthalmic and to mucous membranes including vaginal and rectal delivery), pulmonary, e.g., by inhalation or insufflation of powders or aerosols, including by nebulizer; intratracheal, intranasal,
25   epidermal and transdermal), oral or parenteral. Parenteral administration includes intravenous, intraarterial, subcutaneous, intraperitoneal or intramuscular injection or infusion; or intracranial, e.g., intrathecal or intraventricular, administration.

Pharmaceutical compositions and formulations for topical administration may include transdermal patches, ointments, lotions, creams, gels, drops, suppositories, sprays, liquids and powders. Conventional
30   pharmaceutical carriers, aqueous, powder or oily bases, thickeners and the like may be necessary or desirable.

Compositions and formulations for oral administration include powders or granules, suspensions or solutions in water or non-aqueous media, capsules, sachets or tablets. Thickeners, flavoring agents, diluents, emulsifiers,dispersing aids or binders may be desirable.

Compositions and formulations for parenteral, intrathecal or intraventricular administration may include sterile aqueous solutions which may also contain buffers, diluents and other suitable additives such as, but not limited to, penetration enhancers, carrier compounds and other pharmaceutically acceptable carriers or excipients.

Pharmaceutical compositions of the present invention include, but are not limited to, solutions, emulsions, and liposome-containing formulations. These compositions may be generated from a variety of components that include, but are not limited to, preformed liquids, self-emulsifying solids and self-emulsifying semisolids.

The pharmaceutical formulations of the present invention, which may conveniently be presented in unit dosage form, may be prepared according to conventional techniques well known in the pharmaceutical industry. Such techniques include the step of bringing into association the active ingredients with the pharmaceutical carrier(s) or excipient(s). In general, the formulations are prepared by uniformly and intimately bringing into association the active ingredients with liquid carriers or finely divided solid carriers or both, and then, if necessary, shaping the product.

The compositions of the present invention may be formulated into any of many possible dosage forms such as, but not limited to, tablets, capsules, liquid syrups, soft gels, suppositories, and enemas. The compositions of the present invention may also be formulated as suspensions in aqueous, non-aqueous or mixed media. Aqueous suspensions may further contain substances which increase the viscosity of the suspension including, for example, sodium carboxymethylcellulose, sorbitol and/or dextran. The suspension may also contain stabilizers.

In one embodiment of the present invention, the pharmaceutical compositions may be formulated and used as foams. Pharmaceutical foams include formulations such as, but not limited to, emulsions, microemulsions, creams, jellies and liposomes. While basically similar in nature these formulations vary in the components and the consistency of the final product. The preparation of such compositions and formulations is generally known to those skilled in the pharmaceutical and formulation arts and may be applied to the formulation of the compositions of the present invention.

The formulation of therapeutic compositions and their subsequent administration is believed to be within the skill of those in the art. Dosing is dependent on severity and responsiveness of the disease state to be treated, with the course of treatment lasting from several days to several months, or until a cure is effected or a diminution of the disease state is achieved. Optimal dosing schedules can be calculated from

measurements of drug accumulation in the body of the patient. Persons of ordinary skill can easily determine optimum dosages, dosing methodologies and repetition rates. Optimum dosages may vary depending on the relative potency of individual oligonucleotides, and can generally be estimated based on EC50 as found to be effective in in vitro and in vivo animal models.

5   In general, dosage is from 0.01 ug to 10 g per kg of body weight, and may be given once or more daily, weekly, monthly or yearly, or even once every 2 to 20 years. Persons of ordinary skill in the art can easily estimate repetition rates for dosing based on measured residence times and concentrations of the drug in bodily fluids or tissues. Following successful treatment, it may be desirable to have the patient undergo maintenance therapy to prevent the recurrence of the disease state, wherein the oligonucleotide is

10   administered in maintenance doses, ranging from 0.01 ug to 10 g per kg of body weight, once or more daily, to once every 20 years.

Gene Therapy

         As noted above, SNARE YKT6 is necessary for cell growth, POLD2 is involved in DNA replication and repair, AEBP1 is involved in repressing adipogenesis and glucokinase is involved in

15   glucose sensing in pancreatic islet beta cells and liver. Therefore, the SNARE YKT6 gene may be used to modulate or prevent cell apoptosis and treat such disorders as virus-induced lymphocyte depletion (AIDS); cell death in neurodegenerative disorders characterized by the gradual loss of specific sets of neurons (e.g., Alzheimer's Disease, Parkinson's disease, ALS, retinitis pigmentosa, spinal muscular atrophy and various forms of cerebellar degeneration), cell death in blood cell disorders resulting from deprivation of growth

20   factors (anemia associated with chronic disease, aplastic anemia, chronic neutropenia and myelodysplastic syndromes) and disorders arising out of an acute loss of blood flow (e.g., myocardial infarctions and stroke). The glucokinase gene may be used to treat diabetes mellitus. The AEBP1 gene may be used to modulate or inhibit adipogenesis and treat obesity, diabetes mellitus and/or osteopenic disorders. POLD2 may be used to treat defects in DNA repair such as xeroderma pigmentosum, progeria and ataxia

25   telangiectasia.

         As described herein, the polynucleotide of the present invention may be introduced into a patient's cells for therapeutic uses. As will be discussed in further detail below, cells can be transfected using any appropriate means, including viral vectors, as shown by the example, chemical transfectants, or physico-mechanical methods such as electroporation and direct diffusion of DNA. See, for example, Wolff, Jon A,

30   et al., "Direct gene transfer into mouse muscle in vivo," *Science,* 247, 1465-1468, 1990; and Wolff, Jon A, "Human dystrophin expression in mdx mice after intramuscular injection of DNA constructs," *Nature, 352,* 815-818, 1991. As used herein, vectors are agents that transport the gene into the cell without degradation and include a promoter yielding expression of the gene in the cells into which it is delivered. As will be discussed in further detail below, promoters can be general promoters, yielding expression in a variety of

mammalian cells, or cell specific, or even nuclear versus cytoplasmic specific. These are known to those skilled in the art and can be constructed using standard molecular biology protocols. Vectors have been divided into two classes:

a)   Biological agents derived from viral, bacterial or other sources.

5   b)   Chemical physical methods that increase the potential for gene uptake, directly introduce the gene into the nucleus or target the gene to a cell receptor.

Biological Vectors

Viral vectors have higher transaction (ability to introduce genes) abilities than do most chemical or
10   physical methods to introduce genes into cells. Vectors that may be used in the present invention include viruses, such as adenoviruses, adeno associated virus (AAV), vaccinia, herpesviruses, baculoviruses and retroviruses, bacteriophages, cosmids, plasmids, fungal vectors and other recombination vehicles typically used in the art which have been described for expression in a variety of eukaryotic and prokaryotic hosts, and may be used for gene therapy as well as for simple protein expression. Polynucleotides are inserted into
15   vector genomes using methods well known in the art.

Retroviral vectors are the vectors most commonly used in clinical trials, since they carry a larger genetic payload than other viral vectors. However, they are not useful in non-proliferating cells. Adenovirus vectors are relatively stable and easy to work with, have high titers, and can be delivered in aerosol formulation. Pox viral vectors are large and have several sites for inserting genes, they are
20   thermostable and can be stored at room temperature.

Examples of promoters are SP6, T4, T7, SV40 early promoter, cytomegalovirus (CMV) promoter, mouse mammary tumor virus (MMTV) steroid-inducible promoter, Moloney murine leukemia virus (MMLV) promoter, phosphoglycerate kinase (PGK) promoter, and the like. Alternatively, the promoter may be an endogenous adenovirus promoter, for example the E1 a promoter or the Ad2 major late promoter
25   (MLP). Similarly, those of ordinary skill in the art can construct adenoviral vectors utilizing endogenous or heterologous poly A addition signals.

Plasmids are not integrated into the genome and the vast majority of them are present only from a few weeks to several months, so they are typically very safe. However, they have lower expression levels than retroviruses and since cells have the ability to identify and eventually shut down foreign gene
30   expression, the continuous release of DNA from the polymer to the target cells substantially increases the duration of functional expression while maintaining the benefit of the safety associated with non-viral transfections.

Chemical/physical vectors

Other methods to directly introduce genes into cells or exploit receptors on the surface of cells include the use of liposomes and lipids, ligands for specific cell surface receptors, cell receptors, and calcium phosphate and other chemical mediators, microinjections directly to single cells, electroporation and homologous recombination. Liposomes are commercially available from Gibco BRL, for example, as LIPOFECTIN® and LIPOFECTACE®, which are formed of cationic lipids such as N-[l-(2,3 dioleyloxy)-propyl]-n,n,n-trimethylammonium chloride (DOTMA) and dimethyl dioctadecylammonium bromide (DDAB). Numerous methods are also published for making liposomes, known to those skilled in the art.

For example, Nucleic acid-Lipid Complexes--Lipid carriers can be associated with naked nucleic acids (e.g., plasmid DNA) to facilitate passage through cellular membranes. Cationic, anionic, or neutral lipids can be used for this purpose. However, cationic lipids are preferred because they have been shown to associate better with DNA which, generally, has a negative charge. Cationic lipids have also been shown to mediate intracellular delivery of plasmid DNA (Felgner and Ringold, Nature 337:387 (1989)). Intravenous injection of cationic lipid-plasmid complexes into mice has been shown to result in expression of the DNA in lung (Brigham et al., Am. J. Med. Sci.298:278 (1989)). See also, Osaka et al., J. Pharm. Sci. 85(6):612-618 (1996); San et al., Human Gene Therapy 4:781-788 (1993); Senior et al., Biochemica et Biophysica Acta 1070:173-179 (1991); Kabanov and Kabanov, Bioconjugate Chem. 6:7-20 (1995); Remy et al., Bioconjugate Chem. 5:647-654 (1994); Behr, J-P., Bioconjugate Chem 5:382-389 (1994); Behr et al., Proc. Natl. Acad. Sci., USA 86:6982-6986 (1989); and Wyman et al., Biochem. 36:3008-3017 (1997).

Cationic lipids are known to those of ordinary skill in the art. Representative cationic lipids include those disclosed, for example, in U.S. Pat. No. 5,283,185; and e.g., U.S. Pat. No. 5,767,099. In a preferred embodiment, the cationic lipid is N4 -spermine cholesteryl carbamate (GL-67) disclosed in U.S. Pat. No. 5,767,099. Additional preferred lipids include N4 –spermidine cholestryl carbamate (GL-53) and 1-(N4 - spermind) -2,3-dilaurylglycerol carbamate (GL-89).

The vectors of the invention may be targeted to specific cells by linking a targeting molecule to the vector. A targeting molecule is any agent that is specific for a cell or tissue type of interest, including for example, a ligand, antibody, sugar, receptor, or other binding molecule.

Invention vectors may be delivered to the target cells in a suitable composition, either alone, or complexed, as provided above, comprising the vector and a suitably acceptable carrier. The vector may be delivered to target cells by methods known in the art, for example, intravenous, intramuscular, intranasal, subcutaneous, intubation, lavage, and the like. The vectors may be delivered via *in vivo* or *ex vivo* applications. *In vivo* applications involve the direct administration of an adenoviral vector of the invention formulated into a composition to the cells of an individual. *Ex vivo* applications involve the transfer of the

adenoviral vector directly to harvested autologous cells which are maintained *in vitro*, followed by readministration of the transduced cells to a recipient.

In a specific embodiment, the vector is transfected into antigen-presenting cells. Suitable sources of antigen-presenting cells (APCs) include, but are not limited to, whole cells such as dendritic cells or

5   macrophages; purified MHC class I molecule complexed to ß2-microglobulin and foster antigen-presenting cells. In a specific embodiment, the vectors of the present invention may be introduced into T cells or B cells using methods known in the art (see, for example, Tsokos and Nepom, 2000, J. Clin. Invest. 106:181-183).

The invention described and claimed herein is not to be limited in scope by the specific

10   embodiments herein disclosed, since these embodiments are intended as illustrations of several aspects of the invention. Any equivalent embodiments are intended to be within the scope of this invention. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims.

15   Various references are cited herein, the disclosure of which are incorporated by reference in their entireties.